SENSING ATMOSPHERIC WATER VAPOUR USING THE GLOBAL POSITIONING SYSTEM

A. Z. A. Combrink

Thesis presented for the degree of
DOCTOR OF PHILOSOPHY
in the Faculty of Engineering and the Built Environment
UNIVERSITY OF CAPE TOWN
August 2006

Supervisor:	Prof. C.L. Merry
Co-supervisor:	Dr. W.L. Combrinck

DEDICATION

This work is dedicated to Mariska who supported and encouraged me whilst I was writing this thesis.

ACKNOWLEDGEMENTS

I would like to express my gratitude to the following individuals whose support made this work possible:

- Prof. Charles Merry from the University of Cape Town, the supervisor of the research presented here, who patiently worked on this thesis with me during his sabbatical and continuously provided me with objective and constructive advice.
- Dr. Ludwig Combrinck, programme leader of HartRAO's Space Geodesy Programme and co-supervisor of the work presented in this thesis, who provided me with numerous research opportunities and whose innovative approach to the encountered problems opened many doors.
- Prof. Rui Fernandes from the University of Beira Interior, Portugal who performed the GIPSY processing (cf. Section 6.2) and produced two figures (cf. Figure 19 and Figure 20) presented in this thesis.
- Dr. Machiel Bos from the University of Porto, Portugal who provided a Matlab® implementation for the generation of ARMA covariance matrices (cf. Equation (6.4)) and produced one figure (cf. Figure 21) presented in this thesis.
- Mr. Herrie van Rooy at the University of Johannesburg who provided Matlab® implementations of multiple regression with dummy variables and the sample autocorrelation function (cf. Equations (F.7) and (G.4)).
- Mr. Joel Ondego Botai at HartRAO's Space Geodesy Programme who performed the OCCAM analysis (cf. Section 4.2.2) presented in this thesis.
- Prof. Harm Moraal from the North-West University, Potchefstroom who provided valuable inputs when Appendix B and Appendix C were written.
- Dr. Wolfgang Schlüter (BKG Fundamentalstation Wettzell, Germany) and Prof. Beat Bürki (ETH Zürich, Switzerland) who made water vapour radiometers available for the CONT05 VLBI campaign.
- Ms. Tracey Gill and Mr. Danie Esterhuyse (South African Weather Service) as well as Mr. Richard Wonnacott and Mr. Stephan Koch (Chief Directorate: Surveys and Mapping) who supplied meteorological and GPS data respectively.

I would also like to acknowledge the support received from the National Research Foundation and Inkaba yeAfrica, a joint Germany–South Africa geoscience initiative.

DECLARATION

I, Adriaan Zacharias Albertus Combrink, hereby declare that the work on which this thesis is based, is original (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree at this or any other University.

I empower the University to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

15 August 2006

ABSTRACT

Atmospheric water vapour measurements are of importance to meteorologists, radio astronomers and geodesists. Precipitable water vapour (PWV) is a greenhouse gas to be reckoned with in numerical weather models and climate change studies, it is a nuisance in centimetre-wavelength radio astronomy and introduces range errors in space geodetic techniques.

The propagation time of electromagnetic waves is the principal observable in the Global Positioning System (GPS). Accurate estimates of the delays experienced by the radio signals travelling from the satellites to ground-based receivers are made during the post-processing of GPS observations. In combination with meteorological observations made at the receiver, the estimated delays can be used to determine the amount of integrated precipitable water vapour along the signal path.

In this thesis an overview of the basic GPS principles and components is provided, as well as a derivation, from first physical principles, of the mechanisms contributing to the delay experienced by a radio signal traversing the ionosphere and troposphere. Implementing this theoretical background, PWV and tropospheric delays are estimated and compared to measurements made by other techniques, namely radiosondes, water vapour radiometry and very long baseline interferometry (VLBI). A high degree of correlation is observed in all instances of inter-technique comparison.

The usefulness of GPS-derived slant delays is demonstrated by their ability to reduce VLBI inter-station baseline repeatabilities when they are included in the VLBI analysis. However, this contributed to a higher mean formal baseline error. Furthermore, it shown that GPS-derived slant delay accuracies, when compared to radiometry, can be improved through the stacking of GPS processing residuals to make corrections for the effects of multipath and antenna phase centre variations. A modified residual stacking (MRS) method is proposed, in which data weighting is based on a measured

autocorrelation function; however, in most instances the more complex MRS failed to significantly improve on the corrections made by normal residual stacking.

GPS-derived PWV time-series from thirty South African stations for a four-year period are presented. A four-parameter model was fitted to the time-series to correct for seasonal effects and detect linear trends. It is shown that an autoregressive moving average (ARMA) model is required to estimate realistic trend uncertainties, rather than the white-noise model implicit in standard least-squares analyses. Furthermore, significant trends in PWV were observed in South Africa with the central parts showing a decrease in PWV during the study period, while an increase is observed over the southwest and northeast. These trends coincide with a temperature increase observed over the whole of South Africa for the study period. A hypothesis is presented to explain the different trends, based on the different sources of PWV in different climate areas.

Lastly, vertical earth tide displacements (VETD) measured by gravimetry are compared to the modelled VETD applied during GPS processing. It is shown that mismodelled VETD can contribute significant errors to GPS-derived PWV. A number of methods to mitigate this error are proposed and compared to each other, including a novel technique to accurately measure VETD by GPS.

Key terms

Precipitable water vapour (PWV)
Global Positioning System (GPS)
Tropospheric delay
Earth tide
Multipath

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

AR	Autoregressive
ARMA	Autoregressive moving average
AS	Anti-spoofing
C/A-code	Coarse/acquisition code
CDMA	Code Division Multiple Access
CDSM	Chief Directorate: Surveys and Mapping
COSMIC	Constellation Observing System for Meteorology, Ionosphere and Climate
DoD	Department of Defense
ECEF	Earth-centred Earth-fixed
ETH	Eidgenössische Technische Hochschule (Federal Institute of Technology)
FDMA	Frequency Division Multiple Access
GAMIT	GPS at MIT
GFZ	GeoForschungZentrum (Research Centre for Geosciences)
GGP	Global Geodynamics Project
GIPSY	GPS-Inferred Positioning System
GIS	Geographic information systems
GPS	Global Positioning System
GLONASS	Global Orbiting Navigation Satellite System
GNSS	Global Navigation Satellite System
HartRAO	Hartebeesthoek Radio Astronomy Observatory
IERS	International Earth Rotation and Reference Systems Service
IF	Intermediate frequency
IGS	International GNSS Service
INMARSAT	International Maritime Satellite
I/O	Input/output
ITCZ	Inter-tropic convergence zone
ITRF	International Terrestrial Reference Frame
JPL	Jet Propulsion Laboratory

JPO	Joint Program Office
LCD	Liquid crystal display
LoS	Line-of-sight
MA	Moving average
MAP	Mean annual precipitation
MCS	Master Control Segment
MIT	Massachusetts Institute of Technology
MJD	Modified Julian day
MRS	Modified residual stacking
NASA	National Aeronautics and Space Administration
NAVSTAR	Navigation System with Timing and Ranging
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NMF	Niell mapping function
NOAA	National Oceanic and Atmospheric Administration
NRCan	Natural Resources Canada
OCS	Operational Control Segment
P-code	Precision code
ppm	Parts per million
PPP	Precise Point Positioning
PRN	Pseudorandom noise
PVT	Position, velocity and time
PWV	Precipitable water vapour
RAM	Random-access memory
RF	Radio frequency
RHCP	Right-hand circularly polarised
rms	Root-mean-square
RS	Radiosonde
RS-232	Recommended Standard 232
SA	Selective availability
sACF	Sample autocorrelation function

SADC	Southern African Development Community
SAWS	South African Weather Service
SG	Super-conducting gravimeter
SOPAC	Scripps Orbit and Permanent Array Center
SV	Space vehicle
TEC	Total electron content
TECU	TEC unit
TT&C	Telemetry, tracking and command
VETD	Vertical earth tide displacement
VLBI	Very long baseline interferometry
VTEC	Vertical TEC
WAAS	Wide Area Augmentation Service
WVR	Water vapour radiometer
ZHD	Zenith hydrostatic delay
ZTD	Zenith tropospheric delay
ZWD	Zenith wet delay

1 INTRODUCTION

Atmospheric water vapour is of interest to radio astronomers, space geodesists and meteorologists for various reasons. To the radio astronomer observing electromagnetic waves at the centimetre wavelength level, atmospheric water vapour is a nuisance, causing absorption and emission at these wavelengths. Atmospheric water vapour also causes satellites, used in space geodetic techniques, to appear further than they really are, as a result of the refraction of electromagnetic waves. Accurate measurements of water vapour enable radio astronomers and space geodesists to correct their observations by including it in tropospheric models. Meteorologists bring atmospheric water vapour a little closer to the man in the street. The atmosphere's water vapour, an abundant and highly variable greenhouse gas, is an extremely important parameter in numerical weather prediction.

Radio wave propagation time is the primary observable in the Global Positioning System (GPS) (Kaplan, 1996:18). The recorded phase measurements and local meteorological observations can be used to estimate the amount of atmospheric water vapour, which is of potential use to the above-mentioned communities.

This thesis attempts to address the needs of South African meteorologists and radio astronomers regarding the measurement and usefulness of precipitable water vapour (PWV), as observed using GPS. Of interest to the space geodesy community in general is a number of methods presented here, which attempts to improve the accuracy of tropospheric delay estimates, as derived from GPS observations.

A particular objective of this thesis is to provide a theoretical reference for radio astronomers and meteorologists regarding the derivation of PWV from GPS observations. Chapter 2 provides an overview of the Global Positioning System, while Chapter 3 describes the effects of the ionosphere and troposphere on the propagation of radio signals traversing the atmosphere from the satellite to the ground-based receiver.

Appendix B and Appendix C provide the derivation of the ionospheric delay and the tropospheric index of refraction respectively, from first principles in physics.

In Chapter 4 the GPS-derived PWV is compared to PWV derived from other techniques, namely the South African Weather Service's radiosondes, and water vapour radiometry and very long baseline interferometry (VLBI) as performed at the Hartebeesthoek Radio Astronomy Observatory. The validation of GPS-derived atmospheric delays is extended in Chapter 5, where line-of-sight slant delays and possible accuracy improvement methods are considered.

Of further interest to the meteorology community is the study presented in Chapter 6 of GPS-derived PWV time series for the South African GPS networks and its use in climate and climate change studies.

Lastly, Chapter 7 considers the possible accuracy improvement of GPS-derived PWV by using observations of solid earth tide from super-conducting gravimeters or GPS measurements.

Chapter 8 concludes the thesis, providing a discussion and critical evaluation of the obtained results and suggesting future research.

2 THE GLOBAL POSITIONING SYSTEM (GPS)

2.1 Introduction

The Global Positioning System is a satellite-based radio navigation system, designed and controlled by the Joint Program Office (JPO) located at the U.S. Air Force System Command's Space Division, Los Angeles Air Force Base (Hoffmann-Wellenhof *et al.*, 1993:13). In 1973 JPO was directed by the U.S. Department of Defense (DoD) to develop and deploy a spaceborne positioning system, that offers global coverage, continuous all-weather operation and high accuracy and serves high-dynamic platforms (Kaplan, 1996:2). This system has become known as the Navigation System with Timing and Ranging (NAVSTAR) Global Positioning System (GPS).

2.2 GPS components

Leick (1990:54) states that GPS comprises three segments, namely the "Space Segment" (satellite constellation), the "Operational Control Segment" (ground control and monitoring network) and the "User Equipment Segment" (user receiving equipment). Each of these three segments will be elaborated on in the following sections.

2.2.1 The Space Segment

The GPS constellation nominally consists of 24 satellites, which are uniformly dispersed in six nearly circular, Earth-centred orbital planes, and a number of active spares. The six orbital planes each have an inclination of 55° relative to the equatorial plane and are equally spaced about the equator at a 60° separation. The satellites orbit at an altitude of approximately 20 200 km, corresponding to an orbital radius (*i.e.* nominal

distance from the centre of mass of the Earth to the satellite) of approximately 26 560 km (Jekeli, 2001:257). The satellites have an orbital period of 11.967 hours, or one half of a sidereal day (Cilliers *et al.*, 2003:51). The first GPS satellite was launched in 1978.

2.2.2 The Operational Control Segment (OCS)

OCS is responsible for the tracking and maintenance of satellites in space, monitors satellite health and signal integrity, maintains the orbital configuration of the satellites, updates the satellite clock corrections and ephemerides and comprises three different physical components (Kaplan, 1996:59–66):

- The master control station (MCS) forms the centre of OCS and is located at Falcon Air Force Base, Colorado Springs. The remote monitoring stations passively track the GPS satellites and then pass the raw data, as well as any additional received navigation messages and local weather data, to MCS via the Defense Satellite Communications System or other ground communications systems. MCS processes data from the monitor stations to form clock corrections, ephemeris and almanac data for each satellite. The data are then distributed to uplink facilities for communication to the satellites.
- The monitor stations are responsible for OCS data collection. Each monitor station consists of a dual-frequency GPS receiver that continuously makes range measurements to each satellite, a receiver antenna of which the phase centre is precisely known, two caesium clocks referenced to GPS system time and a U.S. government weather services meteorological station in the vicinity of the monitor station. Meteorological observations are made to make corrections for the tropospheric delay of the received signals, whereas the dual-frequency receivers can be used to make corrections for the ionospheric delay. The monitor stations are located at Colorado Springs, Kwajalein, Ascension Island, Hawaii, Diego Garcia, Cape Canaveral and a few other locations.
- The ground uplink antenna facilities provide the means of commanding and controlling the satellites and uploading data. An uplink facility stores and uploads

telemetry, tracking and command (TT&C) data. A unique TT&C data set, which includes the navigation message, is prepared by MCS for each satellite and stored until the particular satellite is in view at the uplink facility. An S-band data communication uplink is used for the transmission of the data. The locations of uplink facilities were selected to provide maximum satellite coverage; uplink antennas are collocated with the monitor stations at Ascension Island, Kwajalein, Diego Garcia and Cape Canaveral.

2.2.3 The User Equipment Segment

The user's position, velocity and time (PVT) are determined by processing the L-band signals transmitted from the satellites, using receiving equipment (referred to as a GPS receiver).

The evolution of the technology of GPS receivers parallels that of the electronics industry in general. Initial receivers, manufactured in the mid-1970s as part of the concept validation phase, were large, bulky and heavy analogue devices. With today's technology, digital solid-state GPS receivers with comparable or better capability are much smaller and weigh much less, *e.g.* hand-held GPS receivers. Selection of a GPS receiver depends on the user's application and the required accuracy. Kaplan (1996:72–73) lists all the parameters to consider when selecting a GPS receiver.

A GPS receiving set consists of five principal components: antenna, receiver, processor, input/output (I/O) device and power supply (Kaplan, 1996:68).

Satellite signals are received via the antenna, which is right-hand circularly polarised (RHCP) and provides near hemispherical coverage. Typical coverage is 160° with isotropic gain variations from ~2.5 dBi at zenith to ~0.0 dBi at 10° elevation, and usually gain is negative for elevations below 10° (Kaplan, 1996:68–69). Since the satellite signals are RHCP, a conical helix antenna or variation thereof is suitable. The

designs of antennas currently in use vary from helical coils to thin microstrips. GPS receivers and antennae that track satellites on both the L1 and L2 frequencies need to accommodate 20.46 MHz bandwidths on both frequencies, while those tracking satellites on only the L1 frequency must have a bandwidth of at least 2.046 MHz.

Receivers can be divided into two main types: single-frequency (tracking satellites using only the signals emitted on L1) and dual-frequency (tracking satellites using the signals emitted on both L1 and L2). Most receivers have multiple channels whereby each channel tracks the transmission from a single satellite. The received radio frequency (RF) satellite signals are filtered by a passive bandpass prefilter to reduce out-of-band RF interference, followed by a preamplifier (Jekeli, 2001:264). The RF signals are then downconverted to an intermediate frequency (IF), which is sampled and digitised by an analogue-to-digital converter. The samples are forwarded to the digital signal processor, which contains N parallel channels to simultaneously track up to N satellites (where N generally ranges from 5 to 12). Each channel performs tracking measurements and navigation message demodulation, which are forwarded to the processor.

The processor controls and commands the receiver through its operational sequence, including the following steps: channel signal acquisition, signal tracking, data collection and forming the PVT solution. The formulated PVT solution and navigation data are forwarded to the I/O device.

The I/O device is the interface between the GPS receiver and the user. I/O devices are typically of two types: integrated or external. An integrated I/O device usually consists of a control and display unit, *e.g.* a liquid crystal display (LCD) and control panel. Recommended Standard 232 (RS-232) is a common external I/O device for digital data input and output.

The power supply of the receiving equipment can be internal, external or a combination of both. Portable receivers typically use alkaline or lithium batteries, but where an existing power supply is available, converters and regulators can be used to

supply power to the GPS receiver. There usually is an internal battery to maintain data stored in the volatile random-access memory (RAM) and operate a built-in clock, in the event of power disconnection or failure.

2.3 Other satellite navigation systems and augmentations

The Russian Global Orbiting Navigation Satellite System (GLONASS) is an equivalent to GPS (Jekeli, 2001:255–256). As in the case of GPS, the GLONASS programme was started in the mid-1970s to support military navigation requirements. The first GLONASS satellite was launched in 1982 and the system, which is overseen and operated by Russia's Ministry of Defence, was declared briefly operational in 1993 (Kaplan, 1996:439–440). Early system testing convincingly demonstrated that GLONASS could support civilian use while concurrently meeting Russian defence needs, and the programme was consequently broadened to include civilian users.

The space component of the system would eventually consist of a 24-satellite constellation (7 active satellites in each of three orbital planes, with one active spare in each plane). The ground network consists of a number of satellite monitoring stations and data uploading facilities located throughout Russia.

The GLONASS receiving equipment is similar to that of GPS, but even though a few manufacturers produce receivers, which are capable of performing PVT determination using both the GPS and GLONASS signals, the signals and receivers are generally incompatible. This is due to the fact that GPS satellites transmit a unique pseudorandom noise (PRN) code pair on the same frequencies in a code division multiple access (CDMA) format, while each GLONASS satellite transmits on a different frequency pair (*i.e.* employing frequency division multiple access, or FDMA). Initially the frequencies used for L1 ranged between 1602.00 MHz and 1608.75 MHz, and between 1246.00 MHz and 1251.25 MHz for L2. Due to interference with radio

astronomy measurements, the frequency assignments since 2005 are 1598.06 to 1604.25 MHz for L1 and 1242.94 to 1247.75 MHz for L2 (Kaplan, 1996:453).

It is also noteworthy to mention the European equivalent of GPS and GLONASS, which is called Galileo, of which the first satellite was launched in 2005. The space segment of Galileo will consist of a 27-satellite constellation, distributed uniformly in 3 orbital planes with an inclination of 56° relative to the equator (Schueler *et al.*, 2004). Whereas GPS has a number of active spare satellites, Galileo will have three passive spares in the constellation. It is foreseen that the orbital altitude will be 23 222 km. The ground segment will consist of ~40 ground tracking stations, 2 analysis centres and 9 uplink stations.

Galileo will utilise four carrier frequencies to ensure an improved ionospheric correction and add robustness to the system. The carrier frequencies (and their acronyms) are 1.57542 GHz (L1), 1.27578 GHz (E6), 1.20714 GHz (E5b) and 1.17645 GHz (E5a).

GPS, GLONASS and Galileo together form GNSS, or the Global Navigation Satellite System. The research presented here focuses specifically on GPS, since different equipment is needed to receive signals from GLONASS and Galileo satellites. A detailed technical description of GLONASS is given by Kaplan (1996:439–465), while Schueler *et al.* (2004) provide more detail about Galileo.

Some GNSS applications such as precision farming, oil exploration and open pit mining require much better accuracy than standalone GPS or GLONASS. To achieve the higher accuracies, differential techniques are applied, where correlated (common) errors between two or more receivers, performing range measurements to the same satellites, are removed (Jekeli, 2001:278).

Differential techniques require the use of a reference receiver of which the geographical location is precisely known. The difference between the receiver's

surveyed position and its electronically derived position at a discrete time yields the common error. Current practice is to install a network of reference stations throughout a large geographic area, relay their collected data to a central processing station where differential corrections are formed, uplinked to geostationary satellites and broadcast to users. This technique is referred to as wide area differential.

The International Maritime Satellite System (INMARSAT) is an implementation of a wide area differential service. It is an international consortium providing mobile services on a global scale through four geostationary satellites that broadcast differential corrections for both GPS and GLONASS. The accuracy associated with the overlay is a function of numerous factors, most importantly the ground network architecture. Kaplan (1996:467–486) gives a detailed discussion of INMARSAT and the U.S. Federal Aviation Administration Wide Area Augmentation System (WAAS).

Note that differential corrections can also be made during the post-processing of GPS data and are not limited to satellite-based augmentation services.

2.4 Applications of GPS

Both GLONASS and GPS evolved from dedicated military systems to true dual-use. Satellite navigation technology is being utilised in numerous civil and military applications, ranging from leisure hiking to spacecraft guidance. A few applications of GPS are listed here:

- Accurate navigation for the aviation and maritime communities.
- Spacecraft guidance.
- Military targeting and guidance.
- Navigation or tracking of ground-based vehicles.
- Accurate surveying, mapping, datum conversions and geographic information systems (GIS).

- Geodetic applications, such as monitoring tectonic motions, crustal deformation, earth tides and continental break-up.
- Characterisation of the troposphere and ionosphere.
- GPS time transfer.

2.5 GPS satellite signal characteristics

The GPS space vehicles (SVs) transmit coded signals at two different carrier frequencies in the L-band. Denoted by L1 and L2, the frequencies are 1.57542 GHz and 1.22760 GHz respectively (Ros *et al.*, 2000:357). Two frequencies are provided to permit dual-frequency receivers to make corrections for (dispersive) ionospheric delays. Single-frequency users have to estimate ionospheric delays based upon data in the navigation message.

It is worthwhile to note that a third user frequency of 1.17645 GHz, the so-called L5, will be added to GPS in the near-future (Schueler *et al.*, 2004). Signals on this new carrier frequency will be more robust, *i.e.* more resistant to jamming and in combination with L1 and L2, a second-order ionospheric correction can be made.

The carrier frequencies are modulated by spread spectrum codes with a unique PRN sequence associated with each SV and by the navigation data message. Although all SVs transmit at the same two carrier frequencies, it is possible to distinguish between the signals received simultaneously from different satellites because each satellite transmits mutually exclusive PRN codes (Leick, 1990:58–59). The different signals are separated and detected by CDMA.

The L1 frequency ($154 f_0$, with $f_0 = 10.23$ MHz the nominal reference frequency as observed by an observer on the ground) is modulated by two PRN codes (plus navigation data message), the coarse/acquisition code (C/A-code) and the precision code (P-code). The L2 frequency ($120 f_0$) is modulated by only one PRN code at a time. To

compensate for relativistic effects, the output of the SV's frequency standard of 10.23 MHz is offset by $\Delta f/f = -4.45\times10^{-10}$. This results in $\Delta f = -4.55\times10^{-3}$ Hz and $f_0 = 10.22999999545$ MHz (Leick, 1990:59–60).

To the GPS receiver on the ground, the C/A-code has a chipping rate of 1.023×10^{6} chips.s^{-1} ($f_0/10 = 1.023$ MHz) and the P-code has a chipping rate of 10.23×10^{6} chips.s^{-1} ($f_0 = 10.23$ MHz). The term "chip" is used instead of "bit" to convey that no data information is in the PRN codes (Kaplan, 1996:84).

Because GPS was originally designed for military purposes, so-called "selective availability" (SA) was introduced to derate the accuracy of GPS for non-authorised users. SA introduced pseudorandom Doppler errors on the pseudorange measurements and encrypted an offset error into the satellite's broadcast ephemeris and almanac data, causing positional errors for commercial users. However, on the 1st of May 2000 selective availability was de-activated (Ashby, 2002:47), because the commercial value of GPS became apparent. Commercial users of GPS data had also developed signal processing techniques using, among others, Kalman filtering to partially overcome the limitations introduced by SA, and differential GPS had allowed geodesists to mitigate the effects of SA.

Note that a further security feature called "anti-spoofing" (AS), forms part of the GPS design. The possibility of an enemy spoofing a receiver by transmitting a stronger, false precise code that captures and misleads the receiver is significantly reduced by AS through the encryption of the P-code.

2.6 Signal acquisition, satellite tracking and position determination

GPS signals contain, among other things, information about the satellite's approximate position (broadcast ephemeris with an accuracy of ~200 cm (IGS, 2005)), as

well as a time stamp when the signal was emitted. To generate the latter, each GPS satellite carries one or more accurate atomic clocks – also called "atomic time and frequency standards" (King, 2002:1) – for which relativistic corrections are made by an on-board computer.

The principles used to determine the location of a GPS receiver, are Einstein's second postulate, namely that electromagnetic waves in vacuum propagate at the constant speed c, and trilateration. A GPS receiver receives digital signals like these continuously, and can then determine its position. Ranging using time-of-arrival measurements, the used reference coordinate systems and fundamentals of satellite orbits are discussed in detail by Jekeli (2001:3–9, 101–138, 266–269) and are beyond the scope of this thesis.

A GPS receiver must replicate the PRN code that is transmitted by the SV that is being acquired. Then it must shift the phase of the replica code until it correlates with the SV PRN code. When the phase of the GPS receiver replica code matches the phase of the incoming SV code, there is a maximum correlation. When the phase of the replica code is offset by more than one chip on either side of the incoming SV code, there is minimum correlation (Kaplan, 1996:119). Note that the C/A-code has a 1-millisecond period and the P-code has a 7-day period (Kaplan, 1996:39). In this manner, the GPS receiver detects the SV signal in the "code-phase dimension". The GPS receiver must further detect the SV in the "carrier-phase dimension" by replicating the carrier frequency (plus Doppler shift). Thus the GPS receiver locks on to the SV signal phase.

The carrier phase difference can be measured with sufficient precision and results in the instrumental resolution being less than a millimetre in equivalent path length. The dominant source of error in a phase measurement between a single satellite and a single ground station is the unpredictable behaviour of the "time and frequency standards" (*i.e.* clock errors) serving as reference for the transmitter and receiver (King, 2002:1).

Omitting errors other than the clock offset, we determine the satellite-to-user range between a single satellite and single receiver. Let the vector $\mathbf{u}$ be the user

receiver's position with respect to the Earth-centred Earth-fixed (ECEF) coordinate system origin. Let **r** represent the vector offset from the user to the satellite and let **s** be the position of the satellite relative to the coordinate origin.

The satellite-to-user range vector is $\mathbf{r}=\mathbf{s}-\mathbf{u}$ with magnitude $r=\|\mathbf{r}\|=\|\mathbf{s}-\mathbf{u}\|$, as depicted in Figure 1. The distance r is computed by measuring the propagation time required for a satellite-generated ranging code to transit from the satellite to the user receiver antenna.

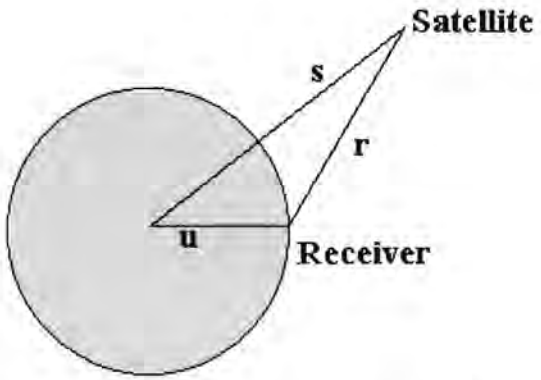

Figure 1: The magnitude *r* of the satellite-to-user range vector is computed from the transit time of the signal from the satellite to the receiver.

Let a specific code, generated by the satellite at T_s, arrive at the receiver at T_u. If the two clocks (satellite and receiver) were perfectly synchronised, the receiver would generate an identical replica code, shift it in time until it achieves correlation with the satellite-generated code, and determine the propagation time $\Delta t = T_u - T_s$. By multiplying this with the speed of light, we obtain the geometric range $r = c(T_u - T_s) = c\Delta t$.

The receiver clock will generally have a bias error t_u from system time, while the onboard atomic clock of the satellite is typically offset δt from system time. Thus, the range determined by the correlation process is denoted as the pseudorange ρ, as it contains the geometric satellite-to-user range, an offset attributed to the difference

between system time and the receiver clock, and an offset between the system time and the satellite clock (Kaplan, 1996:41). Consequently, the pseudorange is $\rho = c\left[(T_u + t_u) - (T_s + \delta t)\right]$. Therefore we can write $\rho - c(t_u - \delta t) = \|\mathbf{s} - \mathbf{u}\|$.

The GPS ground monitoring network determines corrections for the satellite clock offset δt, and the satellite transmits the corrections in the navigation message. The receiver applies these corrections and δt is no longer regarded to be unknown (Kaplan, 1996:43). Hence, the preceding equation can be written as $\rho - ct_u = \|\mathbf{s} - \mathbf{u}\|$.

In order to determine the receiver's position in three dimensions and the offset t_u, pseudorange measurements are made to four satellites resulting in the system of equations $\rho_j = \|\mathbf{s}_j - \mathbf{u}\| + ct_u$, where j ranges from 1 to 4 and references the satellites. These nonlinear equations can be solved by closed form solutions, iterative linearisation techniques or Kalman filtering (Kaplan, 1996:43–44).

The foregoing argument applies to the PVT determination of a single receiver. In the case of a network of receivers, an alternative approach can be used to eliminate errors due to satellite and receiver clock offsets, namely that of double-differencing.

For a single satellite, differencing the phases of signals received simultaneously at each of two ground stations eliminates the effect of satellite clock offset. This measurement is commonly called the single-difference observable. Following the description by Alber *et al.* (2000:2661), let ϕ_1^A and ϕ_2^A be observations of satellites 1 and 2 by receiver A, and ϕ_1^B and ϕ_2^B be observations by receiver B. Single-differences can be obtained by combining these observations: $s_1^{AB} = \phi_1^A - \phi_1^B$ and $s_2^{AB} = \phi_2^A - \phi_2^B$.

A double-difference is formed by differencing the between-station differences between satellites, to completely cancel the effects of variations in the station clocks: $dd_{12}^{AB} = s_1^{AB} - s_2^{AB}$. Since the phase biases of the satellite and receiver oscillators at the

initial epoch are eliminated in double-differenced observations, the double-differenced range is the measured phase plus an integer number of cycles. If the measurement errors – which may arise from mismodelled orbits, receiver noise or a poorly modelled propagation medium – are small compared to a cycle, it is possible to determine the integer values of the biases. An unambiguous measure of double-differenced range is therefore obtained from the initially ambiguous double-differenced phase (Combrink, 2003:4–5). It is noteworthy to mention that over the past two decades several groups have contributed to the improvement of models for the satellite orbits and propagation media, thereby improving the estimation of the integer biases.

2.6.1 The GAMIT suite of GPS processing software

GAMIT (GPS at MIT) is a comprehensive GPS analysis package developed at the Massachusetts Institute of Technology (MIT) and Scripps Orbit and Permanent Array Center (SOPAC) for the estimation of three-dimensional relative positions of ground stations and satellite orbits (King, 2000:1) and is suitable for high-precision estimation of these parameters from GPS surveys. In this section the technique of GPS processing employed by GAMIT is briefly discussed, with specific reference to the elimination of ionospheric effects, modelling of satellite motion and estimation by least squares. GAMIT implements the double-differencing technique, which is also described by King (2002:1–2).

The variable delay introduced by the ionosphere is a major error source in single-frequency GPS measurement. As will be shown in Section 3.1 the ionospheric delay is dispersive; it can furthermore be reduced to less than a millimetre by the formation of the so-called "ionosphere-free linear combination" (LC) of the L1 and L2 phase measurements (King, 2002:3): $\phi_{LC} = 2.546\phi_{L1} - 1.984\phi_{L2}$. Although the formation of LC magnifies the effects of other error sources, GAMIT forms LC to completely eliminate ionospheric effects for networks of extent greater than a few tens of kilometres for which ionospheric errors are uncorrelated.

Therefore, GAMIT incorporates both double-differencing and ionospheric corrections, which require carrier phase observations simultaneously at each epoch from at least two satellites to at least two receivers at both the L1 and L2 frequencies, for the highest relative positioning accuracies (King, 2002:1).

A major requirement of GPS geodetic experiments is the accurate modelling of satellite orbits. Satellite motion is typically described by six parameters (*e.g.* Cartesian position and velocity) and a model off all the forces acting on the satellite during the span of its trajectory. Such a model should contain, amongst other effects, the effects of gravity (from the sun, moon and Earth), solar radiation pressure, and spacecraft gas emission (King, 2002:5). GAMIT generates the trajectory of a satellite by numerical integration of the model's equations of motion.

GAMIT further employs a weighted least squares algorithm to estimate the relative positions of a set of stations, orbital and Earth-rotation parameters, tropospheric zenith delays and phase ambiguities by fitting to double-differenced phase observations (King, 2002:5). The least squares fit for each session is iterated until convergence, since the model relating the observations and parameters is non-linear. The tropospheric zenith delays estimated by GAMIT will be used in this study to obtain precipitable water vapour (PWV) estimates at various GPS receiver sites.

2.7 Influence of the propagation medium on GPS range measurements

After correcting the time-of-arrival of the satellite signals at the receiver by differencing or using the Kalman filter, errors introduced by atmospheric delay still need to be corrected. The atmospheric delay of GPS signals, and the information derived from it, forms the main emphasis of this thesis.

For our purposes the atmosphere is divided into two components, namely the ionosphere (the upper, partially ionised part of the atmosphere) in which the propagation of radio signals are dispersive (*i.e.* the delay of the waves depends on the frequency), and the troposphere (the lower, electrically neutral part of the atmosphere) in which the propagation of radio signals are non-dispersive.

Chapter 3 presents further discussion of atmospheric delay and a detailed derivation thereof from first principles can be found in the appendices.

2.8 Permanent networks of dual-frequency GPS receivers

Several networks of permanent dual-frequency GPS receivers (hereafter referred to as GPS networks) are currently in existence and operational. Apart from being used as satellite tracking networks, these fixed stations can also be used for differential GPS and many geodetic purposes.

Through the establishment of the International GNSS Service (IGS) in 1994 (called the International GPS Service for Geodynamics until 2005), the global scientific community made an effort to promote international standards for GPS data acquisition and analysis, in the light of continued diversification and growth of GPS applications (IGS, 2005). IGS also deployed and operates a common, comprehensive global tracking network. Apart from making GPS observational data available, the products published online in near real-time by IGS to meet numerous scientific and engineering objectives, include:

- GPS satellite ephemerides.
- satellite and station clock offsets and drifts.
- earth rotation parameters, namely polar motion, polar motion rates and length-of-day.
- tropospheric zenith path delay.
- ionospheric total electron content (TEC) grid.

Two GPS networks exist in southern Africa, namely Trignet and the SADC (Southern African Development Community) IGS network. Trignet was set up and is maintained by the South African Chief Directorate: Surveys and Mapping (CDSM). The SADC IGS network is a regional subnetwork of the IGS global tracking network, and was set up by HartRAO's Space Geodesy Programme in collaboration with NASA-JPL (National Aeronautics and Space Administration Jet Propulsion Laboratory), NRCan (Natural Resources Canada), NOAA (National Oceanic and Atmospheric Administration, U.S. Department of Commerce) and the German GeoForschungZentrum (GFZ) Potsdam. The distribution of stations in each of these networks is indicated in Figure 2.

Figure 2: The two existing GPS networks in southern Africa are HartRAO's SADC IGS Network and the Chief Directorate: Surveys and Mapping's Trignet (Cilliers *et al.*, 2003:52).

Data from both these networks are freely available online at http://www.trignet.co.za and ftp://geodesy.hartrao.ac.za/archive/rinex.

3 ATMOSPHERIC EFFECTS

The propagation speed of a wave in a medium can be expressed in terms of the index of refraction n for the medium, which is defined as the ratio of the wave's propagation speed in free space to that in the medium, *i.e.* $n = c/v$, with $c = 2.99792458 \times 10^8$ m.s^{-1}.

A medium is dispersive if its index of refraction is a function of the wave's frequency, so that the propagation velocity v_p of the signal's carrier phase differs from the group velocity v_g associated with the waves carrying the signal information. The information-carrying aspect can be thought of as a group of waves travelling at slightly different frequencies (Jekeli, 2001:271).

The concepts of group and phase velocities and dispersivity are clarified in Appendix A.

Figure 3 shows the different layers (spheres) that make up the earth's atmosphere. For our purposes, we divide the atmosphere into two parts:

- The neutral part, termed the troposphere. It actually includes the troposphere, stratosphere, ozonosphere and the lower part of the mesosphere. It extends over altitudes of 0 – 70 km; and
- The charged part, which is called the ionosphere, made up of the upper mesosphere and the thermosphere, and with altitudes above 70 km.

The refractivity (and the resultant delay) of each of these parts of the atmosphere will be derived in Sections 3.1 and 3.2 (derivations from first principles can be found in Appendix B and Appendix C). It will be shown that the ionosphere is a dispersive medium, and that the troposphere is a non-dispersive medium.

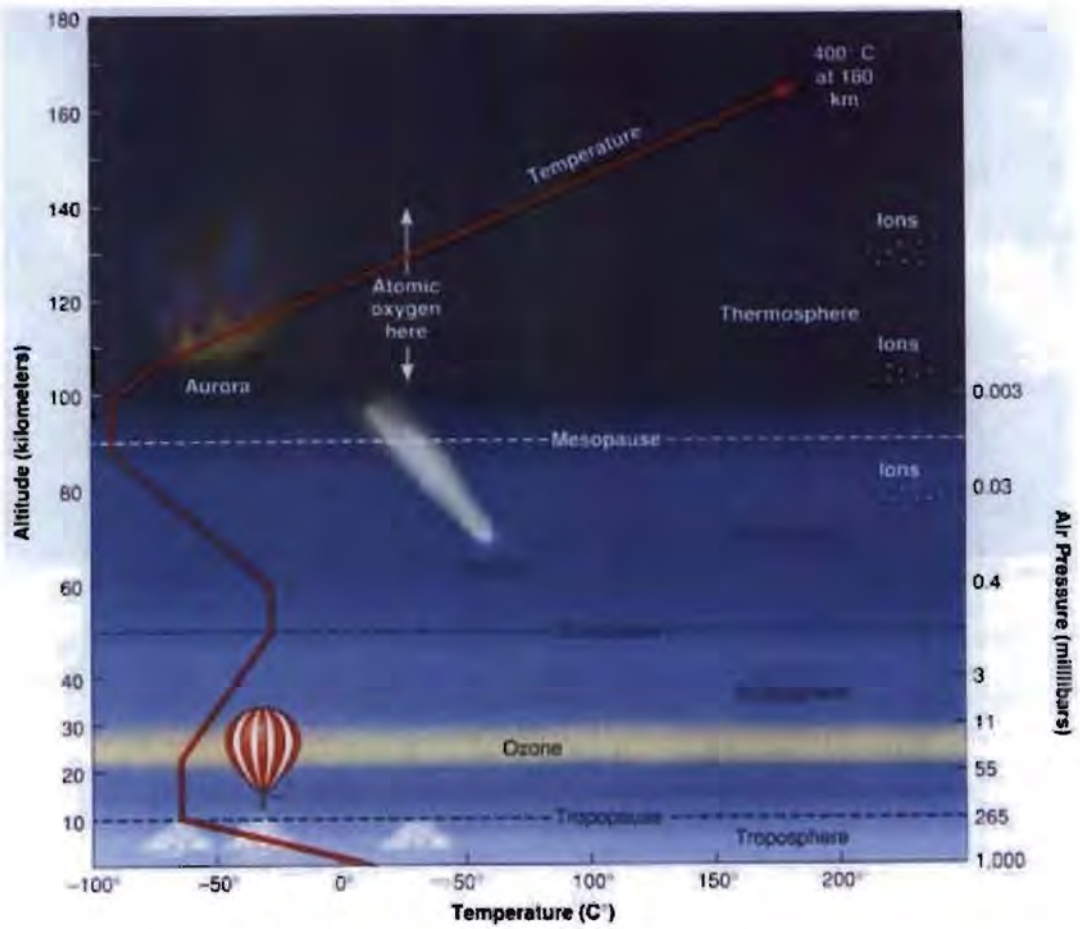

Figure 3: Profile of the earth's atmosphere (Robertson, 2003).

3.1 Ionospheric delay

The ionosphere is the upper, partially charged part of the atmosphere, which is responsible for the absorption and reflection of radio waves at low frequencies (generally below 25 MHz). Atoms in the upper atmosphere are ionised by absorbing radiation energy from the sun – the electrons gain kinetic energy and are displaced from their orbits around the nucleus. Free electrons and ions in the ionosphere continuously recombine, so that only a fraction of the ionosphere is charged at any given time.

The ionosphere has traditionally been split up into three layers, D, E and F, which are classified according to altitude, maximum reflected radio frequency and the most dominant chemical component (Medeiros, 2000). During the daytime, when ionisation from solar radiation is at a maximum, the F layer splits up into two layers, namely the F_1 and F_2 layers. Table 1 gives a summary of the basic characteristics of the ionospheric layers; the altitudes and maximum reflected radio frequencies are averages of quantities that vary with time (these quantities are dependent on season, time of day and solar activity).

Table 1: Properties of the ionospheric layers (Medeiros, 2000).

Layer	Altitude (km)	Maximum reflected frequency (MHz)	Most representative chemical component
D	65 – 80	16	Ozone
E	80 – 115	28	Oxygen
F	140 – 250 (F_1) 250 – 400 (F_2)	16	Nitrogen (F_1) Oxygen (F_2)

The ionosphere is characterised by its content of free electrons and ions. The F_2 layer of the ionosphere has the largest density of charged particles, with values up to $3\times10^{12}\,\mathrm{m}^{-3}$ (Ros *et al.*, 2000:357).

The total electron content (TEC) is defined as the number of electrons in a column of unit area cross section along the transionospheric ray path, written as

$$\mathrm{TEC} = \int_0^{h_0} n_0 dh \tag{3.1}$$

where n_0 is the spatial density of electrons, h is the coordinate along the ray path of the wave, and h_0 corresponds to the effective top of the ionosphere.

TEC is highly variable and depends on several factors, such as local time, geographical location (latitude in particular), season and solar activity. TEC can have

values between 1 TECU (or TEC unit, defined as 10^{16} m^{-2}) and 10^3 TECU (Ros *et al.*, 2000:357).

The delay of radio waves in the ionosphere is derived from first principles in Appendix B. Using the definition of TEC, given in (3.1), the ionospheric delay as derived in (B.17) can be written as

$$\Delta S_{iono} = \mp \frac{40.3 \text{ Hz}^2\text{m}^3}{f^2} \text{TEC} \tag{3.2}$$

where the negative sign applies to phase velocities and the positive sign to group velocities.

Since the ionospheric delay is frequency dependent, it can virtually be eliminated by making range measurements with a dual frequency GPS receiver. Let a pseudorange be presented by

$$\rho_i = r + \Delta S_{iono,i} = r + \frac{c_2}{f_i^2} \tag{3.3}$$

where $i = 1,2$ refers to the GPS L1 and L2 signals, r is the ionosphere-free range and $\Delta S_{iono,i}$ and c_2 are given by (3.2). Recall that (3.3) is a first-order approximation, so that the first-order ionosphere-free range can be given by

$$r = \frac{f_1^2 \rho_1 - f_2^2 \rho_2}{f_1^2 - f_2^2} \tag{3.4}$$

This is referred to as a dual-frequency ionosphere-free linear combination. It is worthwhile to note that it will also be possible in future to form triple-frequency ionosphere-free linear combination. Following Schueler *et al.* (2004), we write the measured pseudorange as

$$\rho_i = r + \Delta S_{iono,i} = r + \frac{c_2}{f_i^2} + \frac{c_3}{f_i^3} \tag{3.5}$$

where $i = 1,2,3$ refers to either GPS L1, L2 and L5 signals, or Galileo L1, E6 and E5(a or b) signals. The second-order ionosphere-free linear combination is then given by

$$r = \frac{f_1^3(f_2 - f_3)\rho_1 + f_1\left(f_3^3\rho_3 - f_2^3\rho_2\right) + f_2 f_3\left(f_2^2\rho_2 - f_3^2\rho_3\right)}{\left[f_1^3 - f_1\left(f_2^2 + f_2 f_3 + f_3^2\right) + f_2 f_3(f_2 + f_3)\right](f_2 - f_3)} \tag{3.6}$$

Schueler *et al.* (2004) have concluded that the "triple-frequency ionosphere-free linear combination has no relevance for practical positioning applications". However, it can be expected that second-order ionospheric corrections will be employed to achieve the highest accuracy requirements of the space geodesy community.

TEC can be determined by differencing the pseudoranges in the first-order approximation. The path length differences ΔS_{iono} of the L1 and L2 signals are given by

$$\Delta S_{iono,L1} = \left(\frac{f_2^2}{f_2^2 - f_1^2}\right)(\rho_1 - \rho_2), \text{ and } \Delta S_{iono,L2} = \left(\frac{f_1}{f_2}\right)^2 \Delta S_{iono,L1} \tag{3.7}$$

The pseudorange differences can further be used to determine TEC along the signal path. From (3.2) and (3.7) follow

$$\text{TEC} = C(\rho_2 - \rho_1) \tag{3.8}$$

defining a new constant

$$C \equiv \frac{1}{40.3\ \text{Hz}^2\text{m}^3}\left(\frac{f_1^2 f_2^2}{f_1^2 - f_2^2}\right) = 5.518 \times 10^{16}\ \text{m}^{-3} \tag{3.9}$$

Since TEC should generally be referenced to the vertical direction through the ionosphere, the slant TEC presented in (3.9) has to be mapped to vertical total electron content (VTEC).

A comparison of ionospheric mapping functions is beyond the scope of this thesis. However, it is appropriate to briefly consider the simple mapping function employed by WAAS (the Wide Area Augmentation System mentioned in Section 2.3). Assuming that all ionospheric ions are concentrated in a single thin layer, the WAAS obliquity factor is given by

$$m(\phi) = \sqrt{1 - \left(\frac{R_e \cos\phi}{R_e + h_I} \right)^2} \tag{3.10}$$

with ϕ the elevation angle of the SV, R_e the radius of the earth and $h_I = 400$ km the height of maximum electron density (Kaplan, 1996:250). (3.8) and (3.10) lead to an expression for VTEC:

$$\text{VTEC} = m \times \text{TEC} = C \sqrt{1 - \left(\frac{R_e \cos\phi}{R_e + h_I} \right)^2} \left[\rho_2 - \rho_1 \right] \tag{3.11}$$

It should be noted that this equation cannot be used directly to compute VTEC because of the differential biases between L1 and L2 pseudoranges, which make its estimation slightly more complicated.

VTEC can be translated back to ionospheric delay using (3.2). A typical value of the VTEC is 50 TECU=5×10^{17} m^{-2}, so that the experienced ionospheric delay of a 1.6 GHz GPS signal in the zenith direction is $\sim \mp 8$m. For a signal arriving at vertical incidence, the delay ranges between 3 m (at night) and 15 m (during the day). At a satellite elevation angle of 10°, the delay can range between 9 m (at night) and 45 m (during the day).

It should be noted that the accuracy of ionospheric corrections determines the accuracy of zenith tropospheric delay (ZTD) estimations. The determination of TEC is therefore relevant to the determination of ZTD and tropospheric precipitable water vapour (PWV) content from GPS data.

3.2 Tropospheric delay

In Appendix C the tropospheric refractive index is derived from first principles and it is shown that, in the radio wavelengths, the troposphere is a non-dispersive medium. In this section we determine the delay of an incoming radio wave in terms of

the apparent length added to the signal's path, and how the dry and wet components of the troposphere contribute to this delay. This will be done in detail, as it forms part of the central theme of this thesis.

Take $\Delta L(\theta)$ to be the geometric delay (or additional apparent path length) of the radio wave, with θ the elevation angle of the direction from which the wave arrives, as shown in Figure 4. Let S be the true path along which the radio wave propagates and G the shortest geometric path along which the signal would traverse if $n=1$. Then

$$\Delta L(\theta) = \int_S n ds - G = \int_S (n-1) ds + (S-G). \tag{3.12}$$

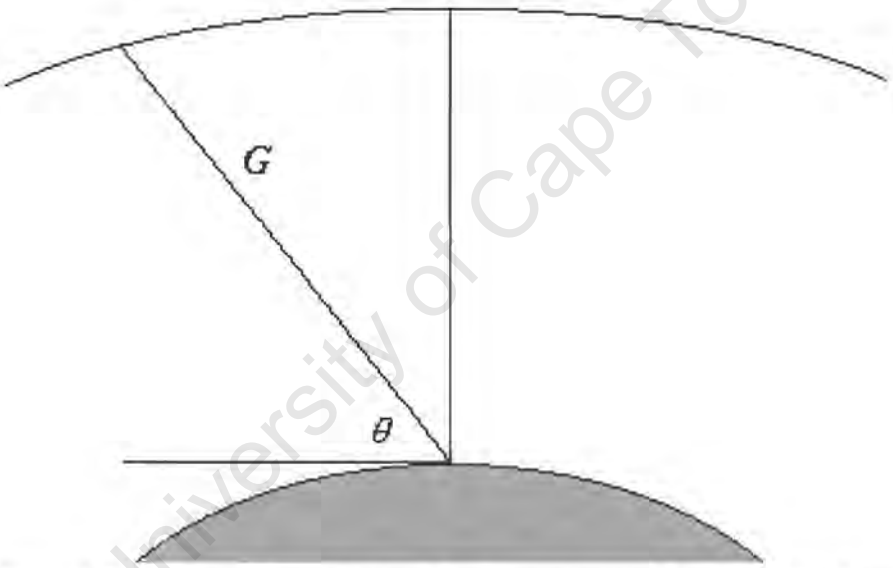

Figure 4: Radio signals from a GPS satellite arrive at a ground-based receiver, from a direction with an associated elevation angle θ above the horizon, travelling along an electrical (true) path; for elevation angles greater than 15 degrees the electrical path can be approximated by the shortest geometric path G.

The first term of (3.12) refers to the delay of a signal due to its reduced speed in the troposphere, while the second term refers to the geometric delay due to the bending of the signal, which can be neglected for elevation angles greater than 15° (Gradinarsky,

2000:12). We can therefore approximate $S = \int_S ds \approx G$. (Note that this is not a valid approximation for lower elevation angles and that the complex bending term should be appropriately corrected for by mapping functions, which are discussed later.)

The tropospheric zenith delay, conventionally abbreviated as ZTD, is one of the quantities estimated by the GPS processing software, and can be defined as $\Delta L^z \equiv \Delta L(90°)$. The focus now shifts to the contributions of water vapour and the "dry" (*i.e.* other than water vapour) gasses to the zenith delay, and we will look at techniques to map the zenith delay to a real measured delay.

In parts per million (ppm) notation, the refractivity can be written as $N = 10^6(n-1)$, so that (3.12) becomes

$$\Delta L = 10^{-6} \int_S N(s)ds. \tag{3.13}$$

We now calculate N and the consequent delay experienced by a radio wave in the troposphere as a function of PWV of the troposphere, and pressure and temperature measured at ground level. This will enable us to estimate PWV from the pressure, the temperature and the delay of a radio wave measured at ground level.

For the troposphere, consisting of, say, q gasses, the delay depends on the density ρ_i of each gas. Following Gradinarsky (2000:12) we write

$$N = \sum_{i=1}^{q} \rho_i (A_i + B_i / T) \tag{3.14}$$

as first derived by Debye (1929:30-35). T is the temperature and A_i and B_i are constants for each gas. In the case of monatomic and diatomic gasses, which account for the biggest part of the atmosphere, $B_i = 0$. However, this is not the case for H_2O, because of its permanent dipole moment. (As defined in Appendix C, the dipole moment of a pair of opposite charges of magnitude q is defined as the magnitude of the charge multiplied by the distance between the charges, and the defined direction is toward the positive

charge.) In his classical work, Debye (1929:30-35) gives a quantitative explanation for the temperature dependence of water's coefficient of refraction. The temperature dependence can also be understood qualitatively, as explained in the following argument (Combrink, 2003:21–22): For higher temperatures, the average distance between atoms in water molecules increases. Therefore, the dipole moment and volume per molecule increase with temperature. Let the temperature dependence of the average distance between the atoms be expressed by $d \propto T^n$, so that the polarisation $P \propto T^{-2n}$, because the dipole moment $p \propto d$ and volume $V \propto d^3$. Consequently, the relative permittivity $\varepsilon_r \propto T^{-2n}$ and the refractivity will decrease as the temperature T increases.

This means that, because the water molecules are permanent electric dipoles, we will have to treat atmospheric water vapour and the "dry" gasses in the atmosphere, separately.

The equation of state for the atmosphere's ith component, with Z_i the dimensionless compressibility, is $p_i V = Z_i n_i R_i T$, where the gas constant of the ith component is defined as $R_i = R/M_i$, with $R = 8.31\ \text{J.K}^{-1}\text{.mole}^{-1}$ the universal gas constant and M_i the molar mass, and all other symbols have their normal meaning. Using $\rho_i = n_i/V$, it is clear that

$$\rho_i = \frac{p_i}{Z_i R_i T}. \tag{3.15}$$

From (3.14) and (3.15) follows that the refractivity is

$$N = \sum_{dry} (A_i p_i / Z_i R_i T) + (A_{wet} p_w / Z_w R_{wet} T) + (B_{wet} p_w / Z_w R_{wet} T^2),$$

where distinction has been made between water vapour and the dry gasses. This can also be expressed as

$$N = k_1 \frac{p_d}{T} Z_d^{-1} + k_2 \frac{p_w}{T} Z_w^{-1} + k_3 \frac{p_w}{T^2} Z_w^{-1}, \tag{3.16}$$

with k_i constants, p_d and p_w the partial pressure of the dry and wet components respectively (Emardson, 1998:5). The first term of (3.16) can be written in terms of the total density, $\rho = \rho_d + \rho_w$. Defining $k_2' = k_2 - k_1 \frac{M_w}{M_d}$, with M_w and $M_d = 28.9644 \pm 0.0014$ g/mol the molar masses of the water vapour and dry air respectively, (3.16) becomes

$$N = k_1 R_d \rho + k_2' \frac{p_w}{T} Z_w^{-1} + k_3 \frac{p_w}{T^2} Z_w^{-1}. \tag{3.17}$$

It is assumed that the condition of hydrostatic equilibrium,

$$\frac{dp}{dz} = -\rho(z) g(z), \tag{3.18}$$

is satisfied, with $g(z)$ the gravitational acceleration and $\rho(z)$ the mass density at a height z, and p the total pressure. The first term of (3.17), which is the hydrostatic component of N, can now be written as

$$N_h = k_1 R_d \rho. \tag{3.19}$$

From (3.13), (3.18) and (3.19) the hydrostatic delay is then expressed as:

$$\Delta L_h = 10^{-6} \int_S N_h ds = 10^{-6} \int_z k_1 R_d \rho(z) dz = 10^{-6} k_1 R_d p_0 g_m^{-1}, \tag{3.20}$$

with p_0 the total pressure and g_m the gravitational acceleration at the centre of mass of the vertical column. According to the model proposed by Saastamoinen (1972:250), $g_m = 9.8062(1 - 0.00266 \times \cos 2\phi - 2.8 \times 10^{-7} H)$ m.s^{-2}, with ϕ the latitude and H the height (in metres) above the geoid.

Davis *et al.* (1985:1603) present the value of $k_1 = 77.604 \pm 0.014$ K/mbar. Substituting this into (3.20), we obtain the hydrostatic delay (in metres)

$$\Delta L_h = \left[(2.2768 \pm 0.0005) \times 10^{-3} \right] \frac{p_0}{f(\phi, H)}, \tag{3.21}$$

with

$$f(\phi, H) = 1 - 2.66 \times 10^{-3} \cos(2\phi) - 2.8 \times 10^{-7} H,$$

where p_0 is measured in millibars and f is dimensionless.. The hydrostatic delay is typically in the range of 1.7 to 2.1 metres; it does not vary much at a specific site and depends mainly on the site's altitude.

The second and third terms of equation (3.17) are the wet components of N, and can be written as

$$N_w = \left[k_2' \frac{p_w}{T} + k_3 \frac{p_w}{T^2} \right] Z_w^{-1}. \tag{3.22}$$

The mean temperature of the vertical column of water vapour is defined as

$$T_m = \int \frac{p_w}{T} dz \Big/ \int \frac{p_w}{T^2} dz. \tag{3.23}$$

From (3.13), (3.22) and (3.23) the delay due to tropospheric water vapour is then expressed as

$$\Delta L_w = 10^{-6} \left[k_2' + \frac{k_3}{T_m} \right] \int_S \frac{p_w}{T} Z_w^{-1} ds. \tag{3.24}$$

Although the hydrostatic delay is the main contributor to the total delay, the wet delay is the main contributor to variation in the total delay, with values ranging between zero and forty centimetres, depending on local meteorological conditions.

The equation of state for water vapour can be written in the form $\frac{p_w}{T} Z_w^{-1} = \frac{nR}{V}$, where all the symbols have their normal meanings. The right hand side can further be written in terms of the specific gas constant R_w and molar mass M_w of water, so that $\frac{p_w}{T} Z_w^{-1} = \frac{nR_w M_w}{V} = D\rho_w R_w$, with ρ_w the density of water (10^3 kg.m^{-3}) and D the relative density of the precipitable water vapour in the troposphere. Defining the total precipitable water vapour content of the troposphere as $\text{PWV} = \int_S D ds$, we can write it as

$$\text{PWV} = \frac{1}{\rho_w R_w} \int_S \frac{p_w}{T} Z_w^{-1} ds = k(T_{surface}) \Delta L_w, \tag{3.25}$$

with $k(T_{surface}) = \dfrac{10^6}{\rho_w R_w \left(k_2' + k_3 T_m^{-1}\right)}$ and the mean temperature approximated as $T_m = 70.2\text{K} + 0.72 T_{surface}$, following Bevis *et al.* (1992:15791). Further, $k_2' = 22.1\text{K.hPa}^{-1}$, $k_3 = 3.739\times10^5\text{K}^2\text{.hPa}^{-1}$ and $R_w = 4.614\times10^2\text{J.K}^{-1}\text{.kg}^{-1}$, according to Borbás (1997:265).

The preceding derivation of (3.25) can be summarised as follows:

- ZTD as estimated during the GPS analysis comprises two components, namely the hydrostatic and the wet zenith delays. This can be written as $\Delta L^z = \Delta L_h + \Delta L_w$.
- The hydrostatic delay can be estimated from local meteorological observations at the GPS site, in particular pressure measurements. By subtracting the hydrostatic component from the total delay, the wet component is obtained.
- The tropospheric water vapour content above a GPS site is directly proportional to the zenith wet delay at the site; the proportionality parameter is a function of the temperature at the site.

We have therefore shown how PWV can be calculated from ZTD if pressure and temperature measurements are available.

However, very few, if any, GPS observations are made at zenith and therefore the discussion now turns to how ZTD can be mapped to estimated delays from arbitrary elevation angles. Conventionally, the tropospheric delay at an arbitrary elevation angle θ is expressed as a function of hydrostatic and wet delays (ΔL_i) and mapping functions ($m_i(\theta)$) respectively:

$$\Delta L(\theta) = \Delta L_h m_h(\theta) + \Delta L_w m_w(\theta) \tag{3.26}$$

Thus, through the use of mapping functions we can map the zenith delay to a delay that a signal would experience at an arbitrary elevation angle.

On the basis of the assumption that the atmosphere in the vicinity of a GPS station is uniform, a simple sine function can be used to map the zenith delay ΔL^z to the delay for an arbitrary elevation angle θ:

$$\Delta L(\theta) = \Delta L^z / \sin(\theta) \tag{3.27}$$

This mapping function is sufficient for elevation angles $\theta > 20°$; for lower elevation angles, more complex functions must be used to account for the curvature of the earth (Gradinarsky, 2000:15).

In Appendix D a brief overview is given of the most commonly encountered empirically derived models for the tropospheric delay, including mapping functions, which can be used successfully for elevation angles below 20°. However, an extensive comparison of the different models and mapping functions is beyond the scope of this thesis and was already performed by Rocken *et al.* (2001:1205–1213) and Janes *et al.* (1991:151–161).

Regarding the use of tropospheric models and mapping functions used for this research project, it is important to note the following:

- The only tropospheric model used, unless otherwise stated, is Saastamoinen's model of the hydrostatic delay component as presented in (3.21). This model suits our purposes very well, as it enables one to isolate the wet delay component (by subtracting the hydrostatic component from the total delay estimated by geodetic techniques) and requires, apart from pressure at the site, no external measurements.
- The Niell mapping function (NMF) was employed in all work done for this research project, unless otherwise stated. The Niell mapping function, proposed by Niell (1996:3230) is most commonly used in space geodesy, because it is presently believed to be accurate at elevation angles from 90° down to below 10°. It does not require any meteorological observations, making it very suitable for our purposes. The typical form of the function is given by (D.1).

4 VERIFICATION OF GPS–DERIVED INTEGRATED WATER VAPOUR

Our attention will now turn to the verification of integrated PWV estimates, as derived from GPS using the GAMIT suite of software described in Section 2.6.1. Water vapour radiometers (WVR), radiosondes (RS) and very long baseline interferometry (VLBI) can be used as independent sources to assess the reliability of GPS-derived PWV (Pacione *et al.*, 2002:309). A brief overview of each of the techniques will be given in the following section. Thereafter, comparisons between RS-derived and GPS-derived PWV from GPS sites collocated with South African Weather Service RS launch sites will be presented. Lastly, GPS-derived ZWD will be compared with the WVR- and VLBI-derived ZWD for the duration of CONT05, a 15-day continuous campaign of VLBI observations from 12 to 27 September 2005.

Comparisons with numerical weather models are omitted. These models typically provide so-called "grid" data, *i.e.* at pre-specified equally spaced coordinates and interpolation could introduce additional errors. Rather, *in situ* measurements, which could be used as input parameters in numerical weather models are considered here.

4.1 Description of independent techniques

4.1.1 RS-derived precipitable water vapour

Precipitable water vapour (PWV) measurements are commonly made using radiosonde launches. Figure 5 is a photograph taken prior to a radiosonde from the South African Weather Service station at Marion Island.

Figure 5: A radiosonde prior to its ascent from the South African Weather Service's station at Marion Island.

The radiosondes measure, amongst other parameters, pressure P, temperature T and relative humidity ρ at altitudes i as it ascends through the atmosphere. The following set of empirical equations can be used to derive PWV (in mm) from the measured parameters:

$$\mathrm{PWV} = \left[Q_{01}(P_0 - P_1) + Q_{12}(P_1 - P_2) + \ldots \right] / g \tag{4.1}$$

where P_i the pressure (in Pa) at i, g the location-dependent gravitational acceleration (in $\mathrm{m.s^{-2}}$), and $Q_{i,i+1}$ the average of the dimensionless specific humidity at i and i+1. The specific humidity is calculated by

$$Q = 0.622 \times \frac{E}{P - 0.378E} \tag{4.2}$$

(USWCL, 1999) where E is the water vapour pressure (in Pa), calculated by

$$E = E_{sat} \times \rho / 100 \tag{4.3}$$

where ρ is relative humidity (in percentage) and

$$E_{sat} = 610.78 \times e^{17.2694T/(238.3+T)} \tag{4.4}$$

is the saturation water vapour (in Pa) at temperature T (in °C) (Kuemmel, 1997).

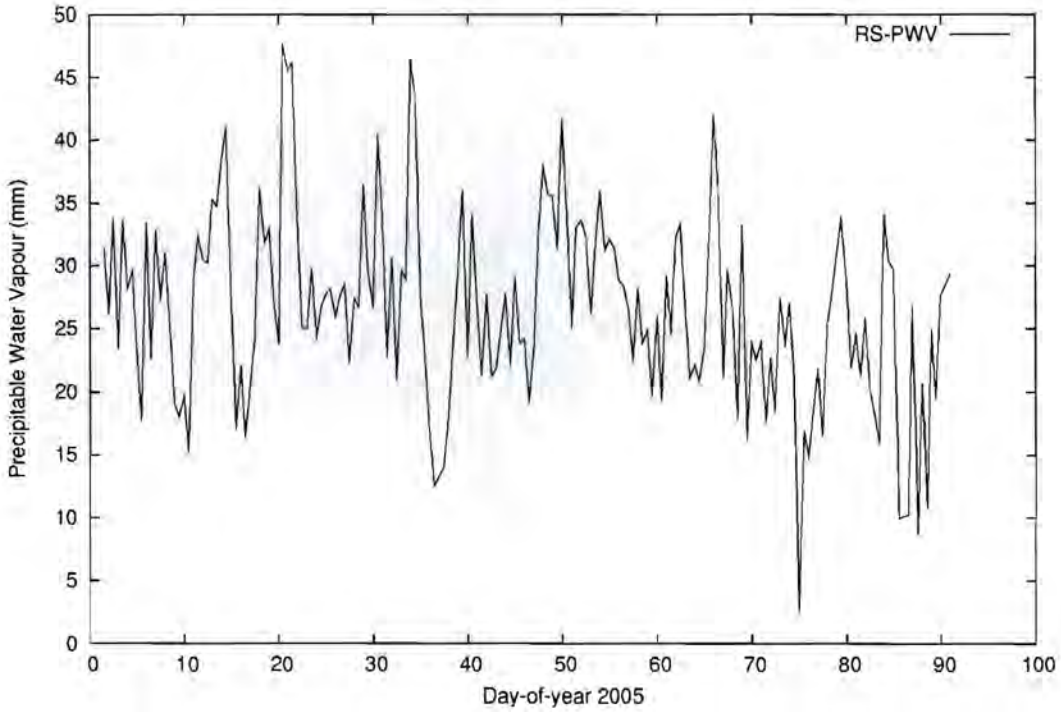

Figure 6: Time-series of RS-derived integrated precipitable water vapour for January to March 2005, at the Port Elizabeth radiosonde launch site.

Software was written to extract integrated PWV as well as radiosonde profiles from raw data, including latitude- and altitude-dependent corrections for g at each sampling altitude i. A time-series of RS-derived PWV is shown in Figure 6 and an example of a water vapour density profile obtained from radiosonde data is shown in Figure 7.

Let the water vapour density at altitude h be represented by $D(h)$. From Figure 7 it is then evident that (4.1) can be rewritten in the following integral form:

$$\mathrm{PWV} = \int_{\text{ground level}}^{\text{top of troposphere}} D(h)\,dh \tag{4.5}$$

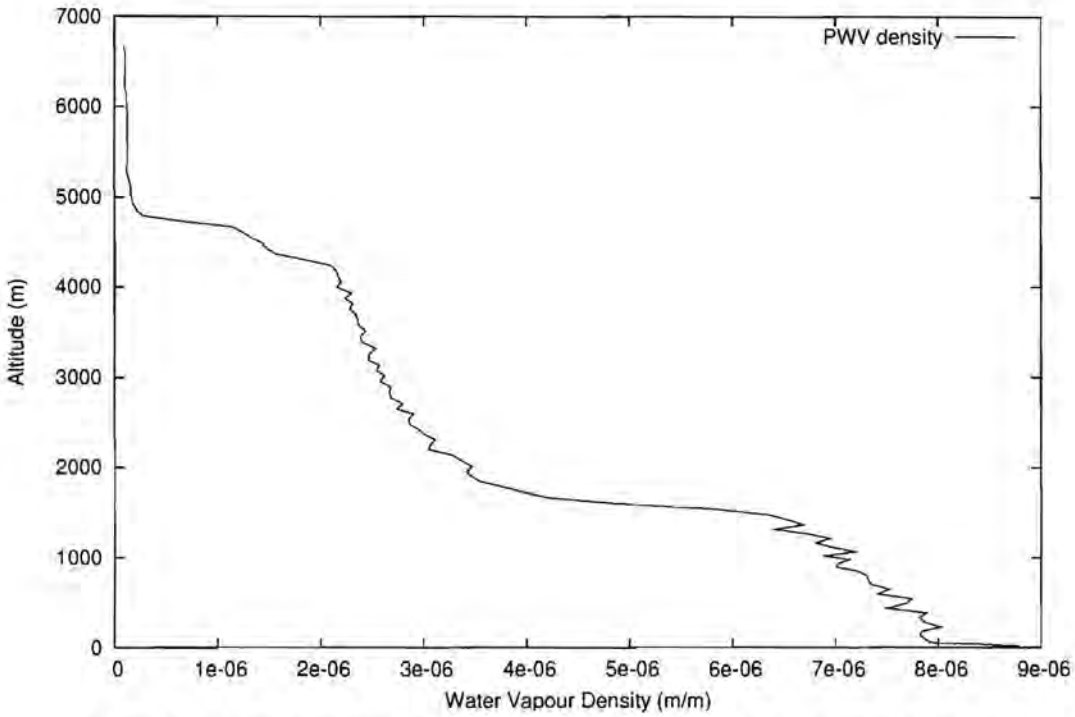

Figure 7: A typical vertical profile of water vapour density, as observed at the De Aar weather station on day 38 of 2005, using a radiosonde.

4.1.2 WVR-derived precipitable water vapour

A water vapour radiometer (WVR), such as the WVR2000 shown in Figure 8, measures the intensity of radiation by water molecules due to thermal excitation (Somieski, 2005:17). In particular, the thermal radiation emitted at the 22.235 GHz characteristic spectral line of water vapour is related to the integrated amount of water vapour in the propagation path of the radiation.

The sky brightness temperature $T_b(f)$ (in Kelvin) observed for a non-scattering atmosphere, is given by the equation of radiative transfer:

Figure 8: A WVR2000 radiometer is operational on the roof of an office building at Hartebeesthoek Radio Astronomy Observatory.

$$T_b(f) = T_c e^{-\tau(f,\infty)} + \int_0^\infty T(z)\alpha(f,z)e^{-\tau(f,z)}dz \tag{4.6}$$

where $T_c = 2.8$ K is the cosmic background temperature, $T(z)$ and $\alpha(f,z)$ are the physical temperature and atmospheric coefficient at height z and frequency f respectively and $\tau(f,z) = \int_0^z \alpha(f,y)dy$ is the opacity (Gradinarsky, 2002:13–14). The first term represents the absorption of incident cosmic background radiation, while the second term represents the atmosphere's emission along the propagation path attenuated by absorption.

In order to make corrections for the influence of liquid water in the atmosphere, thermal radiation should also be measured at a second frequency. In the case of the

WVR2000 (from the Geodesy and Geodynamics Laboratory, Swiss Federal Institute of Technology), as used during this study, observations of sky brightness temperatures are made at the frequencies of 23.8 GHz and 31.5 GHz to observe water vapour and liquid water (Somieski, 2005:17). In the microwave range of 10 to 50 GHz, atmospheric absorption is also due to another major atmospheric constituent, namely oxygen. However, the contribution of oxygen to the total attenuation in the band under consideration is approximately constant (Gradinarsky, 2002:15).

Using the relation $\int_0^\infty \alpha(f,z)e^{-\tau(f,z)}dz = 1-e^{-\tau(f,\infty)}$ and assuming a horizontally stratified isothermal atmosphere with a mean radiating temperature $T_m(f)$, equation (4.6) becomes

$$T_b(f) = T_c e^{-\tau(f,\infty)} + T_m(f)\left[1-e^{-\tau(f,\infty)}\right]$$

(Gradinarsky, 2002:15) and hence the total atmospheric opacity is given by

$$\tau(f,\infty) = \ln\left(\frac{T_m(f)-T_c}{T_m(f)-T_b(f)}\right) \tag{4.7}$$

Note that this mean temperature is different from that defined previously and a linear relation between the ground temperature and the mean radiating temperature has to be estimated from an ensemble of radiosondes (Gradinarsky, 2002:15).

The attenuation (in dB) of the radiation at frequency f is given by

$$A_f = 10\ \log e^{-\tau(f,\infty)} = 4.343\ \tau(f,\infty) \tag{4.8}$$

Gradinarsky (2002:14–16) points out that the total atmospheric attenuation can be expressed as the sum of the atmospheric attenuations due to oxygen, water vapour and liquid water respectively:

$$A_{f,O_2} + A_{f,wv} + A_{f,lw} = \int_0^\infty \left[\alpha_{O_2}(f,z) + \alpha_{wv}(f,z) + \alpha_{lw}(f,z)\right]dz$$

and defines mass absorption coefficients $a_{f,i}$ for water vapour and liquid water as follows:

$$A_{f,wv} = 4.343 \int_0^\infty \alpha_{wv}(f,z)\,dz = a_{f,wv} WV$$

$$A_{f,lw} = 4.343 \int_0^\infty \alpha_{lw}(f,z)\,dz = a_{f,lw} LW$$

These mass absorption coefficients depend on the vertical profiles of temperature, pressure, water vapour and liquid water; average mass absorption coefficients can be estimated from an ensemble of radiosonde profiles (Gradinarsky, 2002:16).

Consequently, in the case of a dual-frequency WVR we arrive at the set of linear equations:

$$A_{f_i} = A_{f_i O_2} + a_{f_i,wv} WV + a_{f_i,lw} LW \text{ with } i = 1,2$$

Assuming that $A_{f_1 O_2} = A_{f_2 O_2}$, we find the amount of water vapour to be

$$WV = b_0 + b_1 A_{f_1} + b_2 A_{f_2} = b_0 + 4.343\left[b_1 \ln\left(\frac{T_m(f_1) - T_c}{T_m(f_1) - T_b(f_1)}\right) + b_2 \ln\left(\frac{T_m(f_2) - T_c}{T_m(f_2) - T_b(f_2)}\right)\right]$$

where the inversion coefficients b_i can once more be determined by radiosonde profiles. Alternatively, for inter-technique comparison wet delays can be calculated by

$$WD = c_0 + c_1 A_{f_1} + c_2 A_{f_2} \tag{4.9}$$

where the inversion coefficients c_i are also determined from radiosonde data; wet delays can be mapped to zenith using standard mapping functions to obtain the zenith wet delay. Similar equations can be derived for the amount of liquid water in the propagation path.

The WVR2000 uses this concept, with the modification of introducing linearised sky brightness temperatures. A detailed description of the implemented retrieval algorithm, the so-called PTX-algorithm, is given by Somieski (2005:19–22). In conclusion, it is important to note that the inversion coefficients, which are determined from radiosonde profiles, have strong seasonal and spatial variation and must therefore be determined accurately for every campaign (*e.g.* a set of inversion coefficients for every month) and for every site.

4.1.3 VLBI-derived precipitable water vapour

During a very long baseline interferometry (VLBI) campaign, large high-gain radio telescopes simultaneously observe distant radio sources in the sky. The 26-metre Hartebeesthoek radio telescope is shown in Figure 9. After the signals from the various locations have been correlated, the relative positions of the antennae can be estimated. Propagation delays of the signal due to the ionosphere and troposphere and clock drifts of the VLBI station's time reference affect the baseline estimates (Gradinarsky, 2002:18). The interested reader is referred to the comprehensive description of VLBI methodologies and results given by Sovers *et al.* (1998:1393–1454).

Figure 9: The 26-metre radio telescope at Hartebeesthoek Radio Astronomy Observatory.

The principal applications of geodetic VLBI are the precise detection of crustal deformation and the estimation of earth orientation parameters and nutation. However, during post-processing tropospheric delays are estimated in order to make the necessary corrections to the baseline estimates. These tropospheric delays can be used to determine the integrated tropospheric PWV by the technique described in Section 3.2.

4.2 Comparison campaigns: methodology and results

4.2.1 Comparison at collocated GPS and radiosonde launch sites

Three months of GPS observational data from South African IGS and Trignet stations were processed using GAMIT 10.2, employing double-differencing and using a ten-degree elevation cut-off. The 24-hour sessions had 30-seconds sampling rates. The Saastamoinen (1972:250) and Niell (1996:3230) models were applied for zenith hydrostatic delay (ZHD) and the mapping functions respectively. Estimates of zenith tropospheric delay (ZTD) were made at 2-hour intervals as a piece-wise linear function. The loosely constrained solution also contained estimates for regional station coordinates, satellite state vectors and phase ambiguities. The IGS final orbits and IERS earth rotation parameters were used. Corrections were applied for the solid earth and polar tides, following the IERS (2003) standards, and for elevation-dependent antenna phase centre variations.

During GPS processing piecewise-linear estimates of the zenith tropospheric delay (ZTD) were made at pre-specified intervals. ZTD estimates at the specified epochs could then be linearly interpolated to the exact time of a radiosonde launch to obtain a ZTD estimate at the specified time. The procedure explained in Section 3.2 was used to obtain a GPS-derived estimate of PWV at the same time as the radiosonde launch, using the pressure and temperature measurements made at ground level at the start of the radiosonde launch. The scripts used to automate this algorithm, also included pressure-corrections for the height-difference between the radiosonde launch site and the phase

centre of the GPS antenna, based on the U.S. Standard Atmosphere 1976 (CRC, 1993:14).

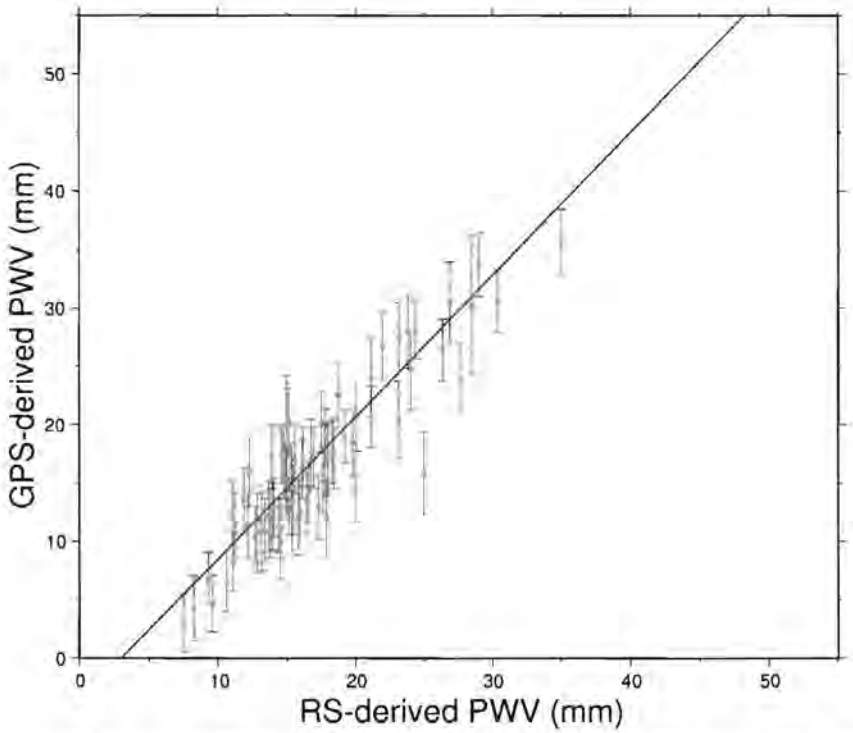

Figure 10: A comparison of the GPS-derived and radiosonde-derived PWV at De Aar, for the period of January to March 2005. The error bars refer to the GPS-derived PWV.

A weighted linear least squares fit was performed on the obtained results to compare the GPS-derived and RS-derived PWV for January to March 2005 at De Aar and Port Elizabeth, as shown in Figure 10 and Figure 11. (Note that these two stations were chosen because they are the only two southern African GPS sites collocated with a year-round radiosonde launch site, excluding Durban. The two stations had also been chosen to include both a relatively dry site, De Aar, and a relatively wet site, Port Elizabeth, in the analysis.)

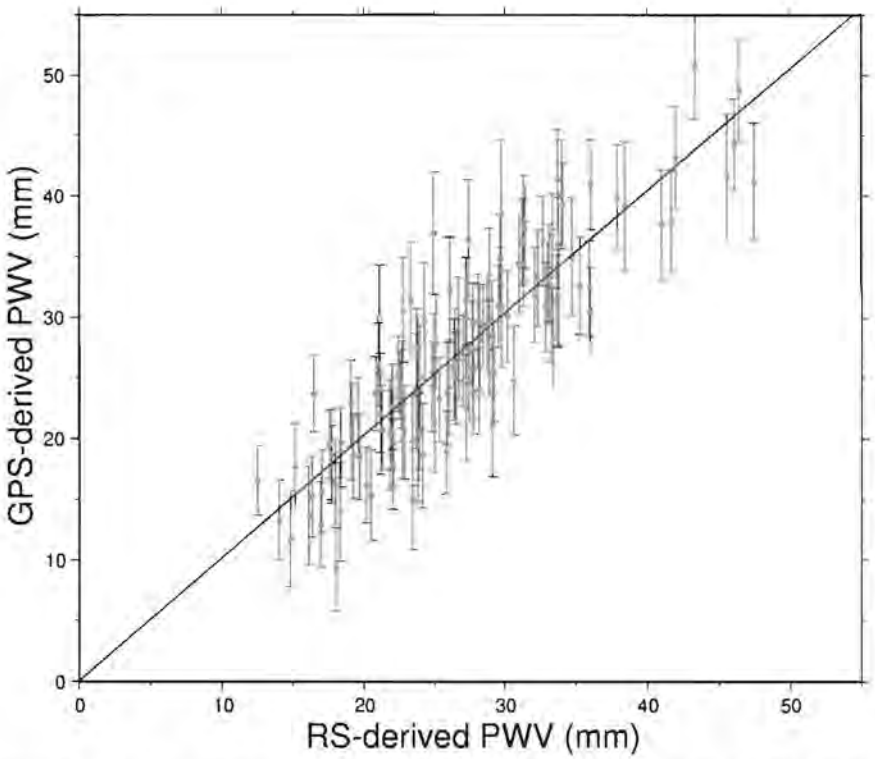

Figure 11: A comparison of the GPS-derived and radiosonde-derived PWV at Port Elizabeth, for the period of January to March 2005. The error bars refer to the GPS-derived PWV.

The obtained correlation coefficients r were 0.91 and 0.88 for De Aar and Port Elizabeth respectively, while the regression coefficients b were 1.18 ± 0.06 and 1.01 ± 0.05 respectively. The respective cut-offs are –3.64 mm and 0.07 mm at RS-PWV = 0. These correspond well to the respective correlation coefficients 0.82 and 0.87 and cut-offs –3.81 mm and 1.20 mm obtained by Wonnacott & Merry (2006) for these two stations, comparing PWV derived from GPS with numerical weather models for March 2004.

A main contributor to the fact that the data do not fit perfectly to a $b = 1$ line, may be erroneous relative humidity measurements (typically the least accurate of the measurements made during the radiosonde's ascent). Wonnacott & Merry (2006) discuss the observed differences between GPS-derived and RS-derived PWV in more detail.

From the high correlation coefficients obtained for the campaigns at these two stations, it is evident that the technique of deriving PWV from GPS observations works equally well in relatively dry and wet conditions. Previously Combrink *et al.* (2004:440,442) found a coefficient of 0.94 for correlation between GPS-derived and RS-derived PWV, during a campaign at De Aar in July 2003. Several other groups, *e.g.* Gradinarsky *et al.* (2002:335–340), also compared GPS-derived PWV with RS-derived PWV, and did not only find the same strong correlation between the two parameters, but also proved conclusively that the technique described in Section 3.2 can be used successfully to measure integrated tropospheric water vapour above a GPS site.

4.2.2 CONT05: Continuous VLBI, GPS and WVR campaign

CONT05 was a two-week campaign of continuous geodetic very long baseline interferometry (VLBI) sessions, from 12 to 27 September 2005. Eleven stations participated in the campaign, including Hartebeesthoek Radio Astronomy Observatory (HartRAO). At HartRAO, continuous GPS and WVR observations were also made simultaneously. Estimates of the zenith wet delay (ZWD) from the three independent, collocated techniques are presented and compared in this section.

A WVR2000 radiometer, built and provided by the Swiss Federal Institute of Technology (ETH Zürich), was operated at HartRAO during the campaign. The required WVR inversion coefficients for HartRAO were derived for the month of September from radiosonde data. The radiosonde observations were made at Irene, approximately 50 km east of HartRAO, by the South African Weather Service (SAWS).

Pottiaux & Warnant (2002:352–353) describe the four basic steps of wet path delay calculation, as utilised by the WVR2000 processing software (version 4.2) used for this campaign. Firstly, during the filtering of raw observations outliers are identified and rejected. Secondly, antenna brightness temperatures are determined by calibrating

measurements to reference signals from the on-board calibration targets, through the so-called "tipping curve analysis." Thirdly, the corrected antenna brightness temperatures are used to compute path delays, as described in Section 4.1.2. Finally, median smoothing and mean smoothing of the zenith wet delay (ZWD) is applied, using a moving window in the time domain.

Fifteen 24-hour sessions of correlated VLBI data were analysed independently using OCCAM version 6.0 (http://ivscc.gsfc.nasa.gov/analysis/occam.html). A five-degree elevation cut-off angle was applied. The Saastamoinen (1972:250) ZHD model was applied and the Niell (1996:3230) dry and wet mapping functions were used. ZWD were estimated hourly as a piece-wise linear function with a $20\,\mathrm{mm}/\sqrt{\mathrm{hour}}$ constraint.

Sixteen days of GPS observational data from South African IGS stations were processed using GAMIT 10.2, employing the same strategy as described in Section 4.2.1 and with the *a priori* hydrostatic delay at HRAO set to be 1.957 m. ZWD for the HRAO GPS station was calculated from the obtained ZTD using pressure measurements from the collocated Paroscientific MET3 unit.

Niell *et al.* (2001:831) point out that precipitable water vapour (PWV) is not uniquely provided by the three techniques under consideration, since the distribution of temperature and water vapour along their ray paths are not known. The delay, or excess electrical path through the atmosphere, is directly estimated by VLBI and GPS, while it is derived by radiometry based on antenna brightness temperatures. Therefore, an additional error would be introduced if the obtained ZWD estimates were converted to PWV. Consequently, we now compare the ZWD as estimated by the three independent techniques and corrected for the height differences between the different instruments.

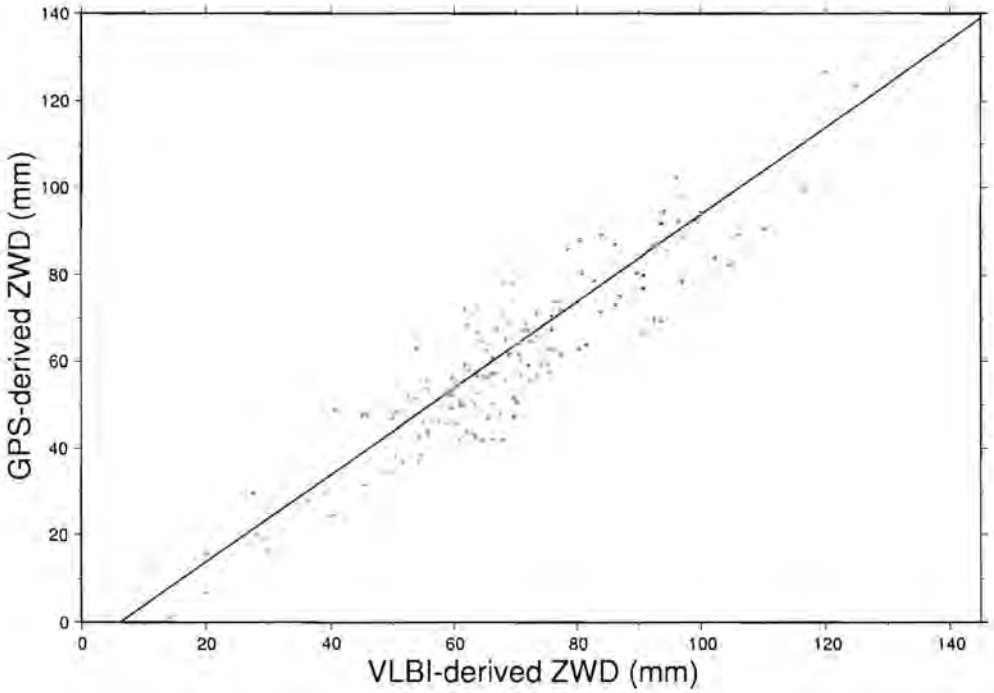

Figure 12: A comparison of the GPS-derived and VLBI-derived ZWD at HartRAO, during the CONT05 campaign (12 to 27 September 2005). The error bars are omitted.

Since the ZWD sampling rates of the three techniques are different, some interpolations had to be performed. GPS processing can support the same estimation rate as VLBI; however, it was found that the most accurate GPS-derived ZTD results are obtained using 2-hourly estimations (King, 1997:4), while the Observatory's routine processing, which was used in this analysis, is set at 1-hourly ZWD estimation. For comparisons of VLBI- and GPS-derived ZWD with WVR data, the six WVR data points before and after the time of ZWD estimation by the geodetic technique were averaged and compared to the ZWD estimated by the other technique. For comparison between the two geodetic techniques, VLBI-derived ZWD was determined as the linear interpolation between the ZWD estimate before and after the time of the GPS estimate; this could be done since the VLBI-derived ZWD was given as a piece-wise linear function.

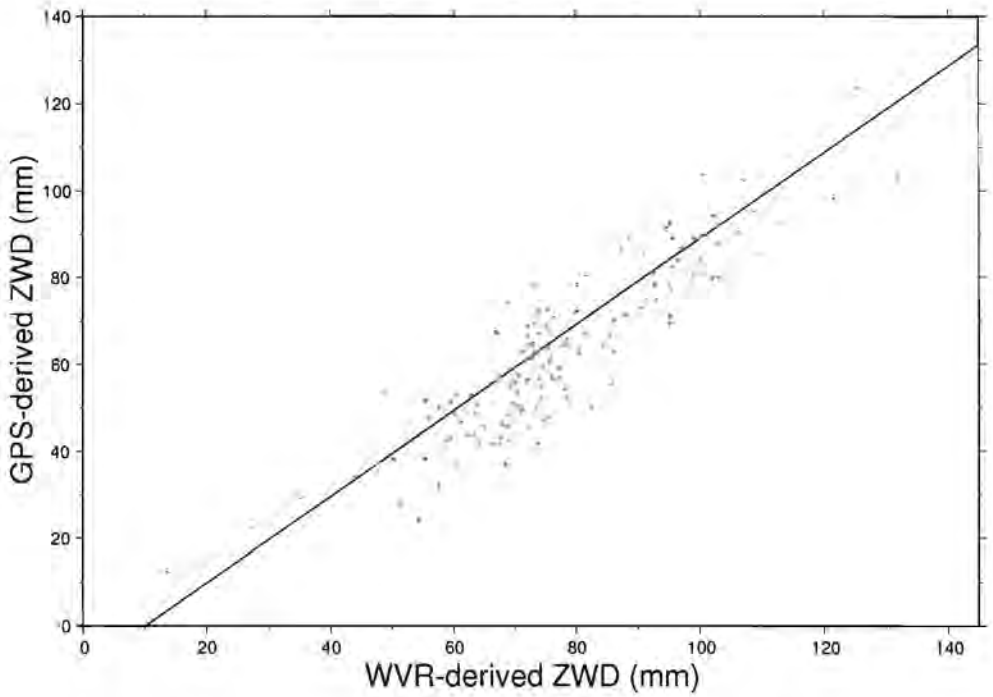

Figure 13: A comparison of the GPS-derived and WVR-derived ZWD at HartRAO, during the CONT05 campaign (12 to 27 September 2005). The error bars are omitted.

ZWD estimates from the three independent techniques are compared to each other in Figure 12 to Figure 14; the comparison is summarised in Table 2, where n is the number of data points compared, r is the correlation coefficient, P is the probability that a random distribution of data would yield a correlation coefficient of magnitude r (see Appendix E), rms is the root-mean-square of the inter-technique residuals (after correcting for technique-dependent offsets), a is the y-axis cut-off of the linear regression and b is the regression coefficient (*i.e.* the slope of the regression line). Data were filtered prior to comparison to remove outliers (less than 5% of the data), tagged as points for which the inter-technique ZWD difference is more than 20 mm (which corresponds to a 2.5-sigma outlier).

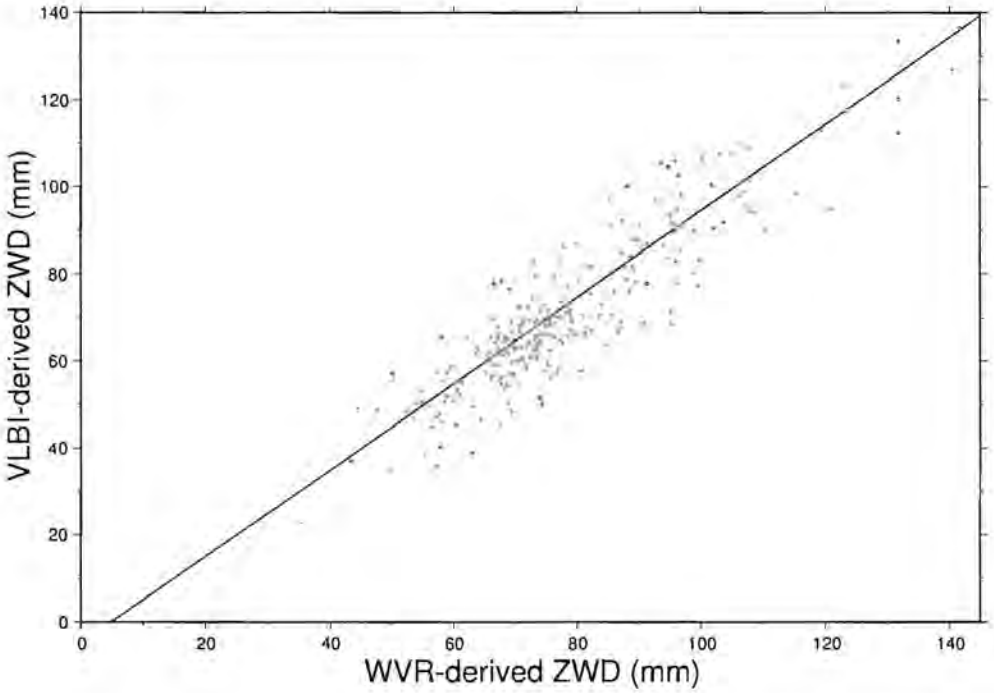

Figure 14: A comparison of the VLBI-derived and WVR-derived ZWD at HartRAO, during the CONT05 campaign (12 to 27 September 2005). The error bars are omitted.

Note that the exact sampling periods of the different techniques differed slightly. GPS sampled continuously during the considered period, while the WVR started operating on the third day of the campaign and the last VLBI experiment ended early on the final campaign day. Consequently, the lowest ZWD values, which were observed on the first and last days of the campaign, were not measured simultaneously by VLBI and WVR.

Note that 8.36 mm ZWD corresponds to ~1.25 mm PWV; currently an accuracy of 1.0 mm PWV or better is required for meteorological applications (Vey *et al.*, 2002:425). Although the biases observed are greater than those observed by Tregoning *et al.* (1998:28701) in their inter-technique comparison of GPS and WVR, the rms differences are slightly smaller. It is evident that ZWD derived by the independent techniques are highly correlated with each other and that the agreement of ZWD

estimates from the different techniques is very nearly at a level acceptable for meteorological applications.

Table 2: Summary of inter-technique ZWD comparison for the CONT05 campaign.

	GPS vs VLBI	GPS vs WVR	VLBI vs WVR
n	152	146	273
r	0.92	0.90	0.90
P	< 10e-10	< 10e-10	< 10e-10
rms (mm)	8.44	8.36	7.75
***a* (mm)**	-6.17	-10.04	-4.85
b	0.97	0.94	0.97

It should be noted that the three techniques compared in this study are all highly sensitive to meteorological observations. The geodetic techniques, GPS and VLBI, rely on accurate measurements of atmospheric pressure at ground-level; a 0.5 hPa error in the pressure would result in an additional error in GPS-derived ZWD of 1.2 mm (Bai & Feng, 2003:85). In the case of radiometry, during meteorological conditions for which the tropospheric water vapour profile differs significantly from the monthly mean, ZWD estimate errors may exceed 1 cm. Errors in emission coefficients can also contribute to significant errors during WVR measurements of ZWD.

The observed technique-dependent biases exist because of the possible presence of seasonal effects and different modelling techniques used to derive ZWD from meteorological and radio frequency observations, *e.g.* mapping functions used, the modelling of hydrostatic delay *etc.* Once long ZWD time-series of collocated instruments are available, the techniques can be better calibrated against each other and the biases can probably be minimised.

Niell *et al.* (2001:830–850) compared ZWD as derived from collocated GPS, VLBI, WVR and radiosondes at Westford, Massachusetts. A number of their conclusions, valid also for our current study, should be highlighted:

- Choosing WVR as the reference measurement (as is done in many studies) does not indicate a judgement as to which instrument provides the most accurate ZWD estimates. Rather, the evaluated techniques are complementary and should be used together to evaluate the accuracies obtained by any technique. A WVR measures a different quantity (integrated brightness temperature) than the geodetic techniques (integrated refractivity) and therefore an improvement of the inter-technique agreement will indicate a better understanding of the troposphere's physical state.
- The observed scale factor between the WVR-ZWD and the other techniques could be expected since the uncertainty of the WVR results is proportional to the amount of water vapour detected.
- A primary limitation of this study is its brief duration. It is likely that seasonal effects influence the independent techniques, which cannot be corrected for given such a short data span.
- It should be possible to improve the accuracy of all the techniques by improving current analysis procedures, applying all known corrections and implementing new models.

5 GEODETIC APPLICATION AND ACCURACY IMPROVEMENTS OF SLANT DELAYS

Estimates of zenith delays are based on observations made in various skyward directions, as described in Section 3.2, and the zenith delays obtained by various techniques were compared to each other in Section 4.2. However, in many regards slant delays are more useful than zenith delays – line-of-sight (LoS) slant delays between a single satellite-receiver pair can be used to:

- improve the repeatability of inter-station baselines as measured by other geodetic techniques;
- observe horizontal distributions of water vapour above a single GPS site, as could be done with water vapour radiometers or thermal infrared sky imagers; and
- perform tomography of the troposphere above a very dense network of GPS receivers.

In this chapter the discussion turns to obtaining slant delays along the line-of-sight between a single GPS satellite and a single GPS receiver from double differences, from which slant wet delays and line-of-sight integrated water vapour estimates can consequently be derived.

Several research groups are actively investigating tomographic techniques, with reasonable success (*e.g.* Flores *et al.*, 2001:439–447). It should however be kept in mind that the GPS networks in southern Africa are not dense enough to be used for tomography of the troposphere, although its density is sufficient for ionospheric tomography. We therefore rather focus our attention on the geodetic application and methods of improving the accuracy of LoS delay estimates, in particular by correcting for the effects of multipath and antenna phase centre variations.

In Section 5.1 baseline results are presented from the CONT05 campaign and it is illustrated how slant delays can be used to improve VLBI inter-station baseline

repeatability. With this application as motivation, methods of improving LoS delay estimates are considered in Section 5.2.

5.1 Geodetic application of GPS-derived slant delays

Given the level of accuracy required when computing geodetic products, atmospheric delay corrections draw the attention of space geodesists. Results are presented here of the repeatability of the baseline between two VLBI stations, namely Hartebeesthoek Radio Astronomy Observatory, South Africa and Fundamentalstation Wettzell, Germany. Both these VLBI stations are collocated with GPS receivers and meteorological sensors.

The daily baselines were first estimated using the CONT05 VLBI data. Thereafter, the baselines were re-estimated by including the GPS-derived slant delays in the VLBI data.

Slant wet delays were calculated for each VLBI observation by mapping the GPS-derived zenith wet delay to the observation's elevation angle, using the Niell (1996:3230) mapping function. The GPS-derived zenith tropospheric delays (ZTD) used for this purpose, are those from the so-called "IGS combined final tropospheric products," which are available from ftp://igs.ensg.ign.fr/pub/igs/tropo/. Zenith wet delays were obtained from the ZTD and pressure measurements, following the description of the technique in Section 3.2. The VLBI-derived slant wet delays for each observation were then replaced in the data files by the GPS-derived slant wet delays, before reprocessing the VLBI data using the OCCAM software and the technique described in Section 4.2.2.

The method described by Kondo *et al.* (1998:1048–1049) is used to calculate the evolution of baseline repeatability during the campaign in order to compare the baselines obtained by the two approaches, of which the latter is referred to as "external" since external (GPS) data were included in the processing. The baseline measurements were

grouped together by a 5-day sliding window. The standard deviation of the baseline and the mean of the formal error were calculated for each group.

The obtained baseline repeatabilities from the two approaches are compared in Figure 15. The mean baseline repeatability obtained in the direct approach was 10.0 mm, while it was 9.8 mm for the external approach.

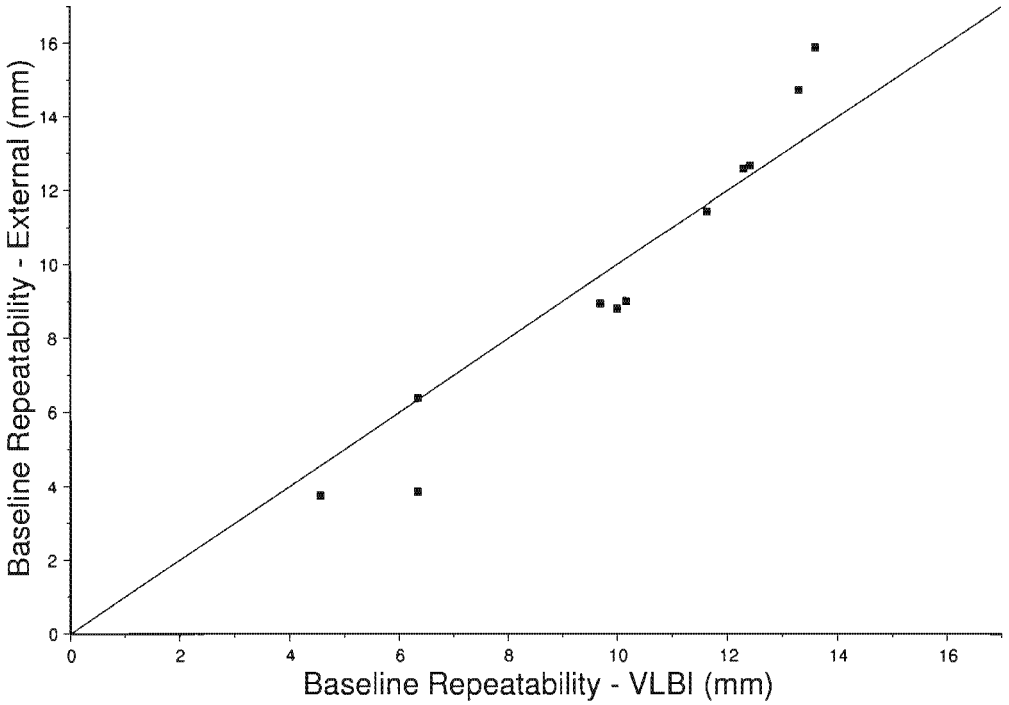

Figure 15: A comparison of the baseline repeatabilities obtained using two different approaches, namely the direct (VLBI) and external (GPS) approaches. The solid line has a 1:1 slope.

The mean formal errors in the direct and external approaches were 3.8 mm and 4.1 mm respectively. This is an expected result, since the formal uncertainties of the GPS-derived delays are greater than those derived from VLBI. The baseline repeatabilities are presented as a function of the mean formal errors in Figure 16.

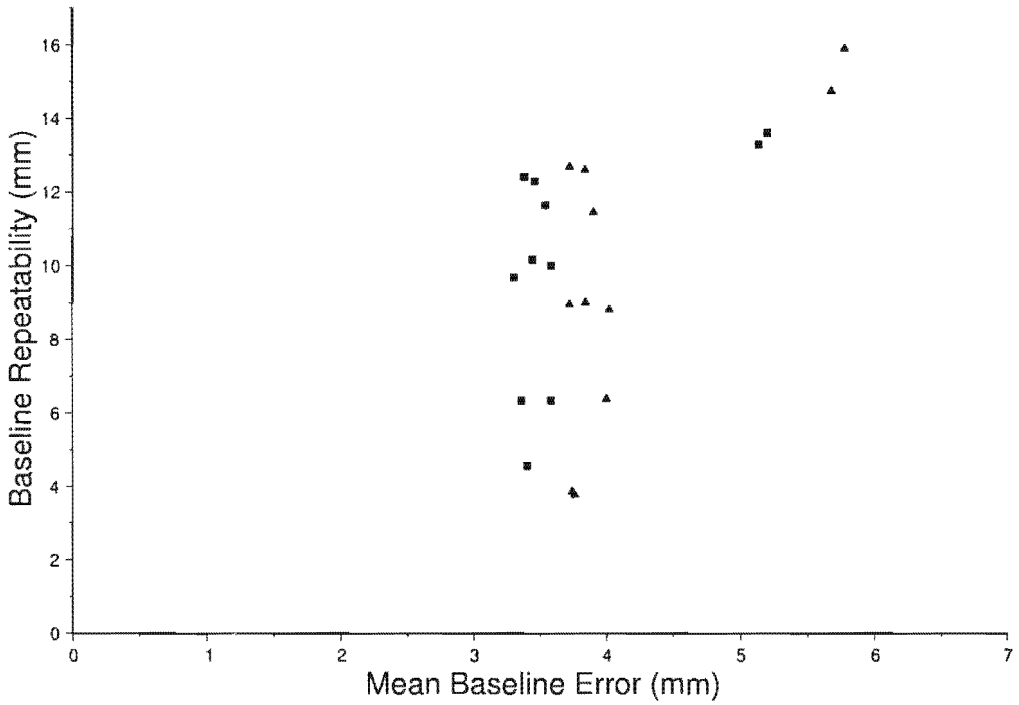

Figure 16: Baseline repeatabilities as a function of the mean formal error. The direct approach is represented by squares, while the triangles represent the external method.

Although the inclusion of GPS-derived delays into the analysis increased the mean formal baseline errors by 10%, it led to an 8% improvement in the baseline repeatabilities. However, this improvement is only at the 65% significance level.

Note that this result was achieved by obtaining GPS slant delays from zenith delays, which were simply mapped to the elevation angle of interest. The rest of this chapter will be dedicated to attempts to improve the accuracy of slant delays, in particular by applying corrections for the effect of multipath.

5.2 Improving slant delay estimate accuracy

5.2.1 LoS delays from GPS observations

We begin by writing the model for total slant delay ΔL, which was first defined in (3.12), as a function of elevation angle ε and azimuth ϕ, as used in GPS processing:

$$\Delta L(\varepsilon,\phi) = \Delta L_w^z \cdot m_w(\varepsilon) + \Delta L_h^z \cdot m_h(\varepsilon) + \frac{dm_w(\varepsilon)}{d\varepsilon}\left[\Xi_n \cos\phi + \Xi_e \sin\phi\right] \tag{5.1}$$

where Ξ_n and Ξ_e are the north-south and east-west gradient parameters, and ΔL_i^z and m_i are the zenith delays and mapping functions for the wet and hydrostatic components respectively (Gradinarsky, 2002:83). In (5.1) it is assumed that the gradient contribution is purely due to water vapour.

The wet slant delay ΔL_w as a function of elevation and azimuth is then given by

$$\Delta L_w(\varepsilon,\phi) = \Delta L_w^z \cdot m_w(\varepsilon) + \frac{dm_w(\varepsilon)}{d\varepsilon}\left[\Xi_n \cos\phi + \Xi_e \sin\phi\right] + \Re(\varepsilon,\phi) \tag{5.2}$$

where $\Re$ is the processing residual, the GPS estimates of the total zenith delay and the gradient parameters are used and the hydrostatic zenith delay is estimated from local pressure measurements (Gradinarsky, 2002:83–84). Recall that this wet delay (cf. (3.24)) is obtained by simply subtracting the hydrostatic delay (cf. (3.21)) from the total delay ΔL.

Gradinarsky (2002:84) defines the last term of (5.2) as the processing residual which contains the model error of the receiver clock, multipath, antenna phase centre variations and the atmospheric variability. The impact of the first three effects must be minimised if we wish for $\Re(\varepsilon,\phi)$ to only contain the atmospheric variability.

The effects of common satellite and receiver clock errors are successfully eliminated by the double-differencing processing method, as employed by the GAMIT

software used. Alber *et al.* (2000:2661–2664) give an extensive description of how the double-differencing technique can be used to eliminate clock errors in order to obtain single path phase delays.

As will be shown in Section 5.2.2, the effect of multipath between a satellite and receiver is a function of the satellite's azimuth and elevation relative to the receiver. Furthermore, the antenna phase centre variation is also a function of the elevation angle of the satellite relative to the antenna. Therefore these noise sources (multipath and phase centre variations) are, to first order, a function of ε and ϕ, and repeat on a sidereal-daily basis according to the direction of the incident ray path. Braun *et al.* (2001:462) described a method to suppress the effects of multipath and phase centre variations by stacking processing residuals together: line-of-sight residuals for a number of days were averaged to produce a so-called "multipath map", which was incorporated into observations, with modest success.

An important conclusion reached by Braun *et al.* (2001:469) is that, applying the techniques of double-differencing and residual stacking, $\Re$ will only contain the true atmospheric variability during conditions and at locations where the non-isotropic water vapour structure is larger than the noise level of the observations (*i.e.* the GPS noise levels).

5.2.2 Multipath

Multipath is defined as the phenomenon whereby a signal arrives at a receiver via multiple paths attributable to reflection and diffraction (Braasch, 1995:547), resulting in a relative phase offset between different radio waves that arrive at a receiver (Bosloper, 1990:6).

In GPS, multipathing distorts the signal's carrier phase (Bosloper, 1990:1). It is a major error source in GPS and degrades the accuracy of interferometric systems.

Furthermore, because GPS employs pseudorange measurements for initialisation purposes, multipath contamination of the pseudorange can increase the time required for initialisation (Braasch, 1995:547).

In this section the phenomenon of multipath will be discussed briefly, by looking at how multipath influences the PRN modulated signal and discussing the sources of multipath.

Influences of multipath on PRN modulated signals

The signal broadcast from a satellite in a PRN ranging system can be written as

$$s(t) = A\cos\left[\omega t + p(t)\frac{\pi}{2}\right] \tag{5.3}$$

where $0.5\times A^2$ is the average power of the received signal, ω is the frequency of the received signal (*i.e.* carrier frequency plus Doppler shift) in radians per second, $p(t)$ is the PRN code which is either -1 or $+1$ (Braasch, 1995:549) and the phase is set to be zero at $t = 0$. Applying trigonometric identities, (5.3) can be reduced to

$$s(t) = -Ap(t)\sin(\omega t) \tag{5.4}$$

The effect of multipath is characterised by its amplitude, time delay and phase, which all are relative to the direct signal. According to Braasch (1995:549), the relative phase of the multipath is a function of the relative time delay and the reflection coefficient of the reflecting object. If the received signal is composed of the direct signal plus a single multipath ray, it can be expressed as

$$s_m(t) = s(t) + \alpha s(t+\delta) \tag{5.5}$$

where α is the multipath relative amplitude and δ is the multipath relative time delay. Following the derivation of Braasch (1995:549), (5.4) can be substituted into (5.5) to yield

$$s_m(t) = -Ap(t)\sin(\omega t) - \alpha Ap(t+\delta)\sin(\omega t + \theta_m) \tag{5.6}$$

where θ_m is the multipath relative phase.

From equation (5.6) the effects of multipath are clear: it distorts the phase of the carrier and could, depending on the phase, attenuate the received signal strength, thereby degrading the signal and introducing errors. The multipath error is thus caused by the interference between the direct and the reflected signals.

Sources of multipath

Multipath can be reduced by reducing the beam width of the transmitting antenna or by increasing the directivity of the receiving antenna, *e.g.* through the use of electronically steered phased arrays. These solutions cannot be implemented in GPS, as the transmitter has to broadcast to many receivers and the receiver has to acquire signals from various satellites (Bosloper, 1990:6). An increase in carrier frequency will also reduce the number of reflective surfaces, but will make the signals susceptible to atmospheric absorption. It seems, therefore, that the multipath error in GPS is here to stay and needs to be dealt with.

Multipath is caused mainly by reflections at the satellite or around the receiver, but also to a lesser extent by multiple signal paths through media such as trees in the vicinity of a GPS antenna. We will consider two kinds of reflection multipath: satellite multipath and receiver multipath.

Satellite multipath occurs when the transmitted signal bounces off from a part of the satellite, before being directed to the earth; its signal strength is about 1% of the total transmitted power (Bosloper, 1990:7). An important feature of the reflected signals is the change in the polarisation from right-hand circularly polarised (RHCP) to left-hand elliptically polarised. An RHCP antenna at the receiver will partially reject the reflected signal, thereby reducing its strength.

With radio waves from a satellite, reflections occur simultaneously at a number of reflective surfaces; each GPS receiver site will have its own unique environment of such reflective surfaces, which are entirely decoupled from the antenna (Bosloper, 1990:6,8).

Reflection changes the magnitude and phase of the reflected wave depending on factors such as reflection coefficient, wave polarisation and the shape of the interface. It can thus be specular (mirror-like) or diffuse (image not retained) depending on the nature of the reflector surface.

The effect of the reflectivity on the received signal depends further on the geometry of the transmitter, receiver and reflector. For each GPS site, the positions of the receiver and reflectors remain fixed. The positions of GPS satellites relative to the GPS sites will be repeated every sidereal day, thereby making multipath a periodic error. Not that this statement is only valid to first order, since reflection coefficients could change over time due to effects such as varying moisture content of some reflectors.

Multipath errors introduced in the C/A-code observables are typically in the order of metres, while multipath errors in the phase observations, such as used in geodetic applications, are in the order of centimetres.

Three main methods exist to mitigate and reduce the effects of multipath in GPS, namely antenna design, software techniques and, most importantly, site selection (EOM, 1998). Multipath can be reduced by simply choosing a good antenna location: a site with a clear view of the sky and with minimal obstructions surrounding it. The design of a GPS antenna for geodetic purposes should also fulfil at least three requirements: reject non-RHCP signals (*i.e.* reflected signals), have a ground plane to prevent signals reflected from the ground to reach the antenna itself and to have a choke ring to reduce the effects of signals reflected off nearby surfaces. Finally, software implementations can detect the presence of a reflected signal, but this (signal processing) technique is most effective when the reflected signal path is tens of metres longer than the direct signal path.

5.2.3 Consecutive-day correlations for the effects of multipath and antenna phase centre variations

The so-called method of "residual stacking," as employed by Braun *et al.* (2001:459–472), simply averages the processing residuals $\Re$ of several days and uses these averages to generate a multipath map. This implies that the same weight is given in the analysis to all the data. A similar approach to calculating the averages would be to use a model employing un-weighted multiple regression with dummy variables as described in equation (F.6) in Appendix F, yielding least-squares estimates rather than averages.

However, one would not necessarily expect the residuals of two days, separated by any number of days between them, to be highly correlated.

To justify this statement, the following arguments should be considered:

- The satellite orbits are repeated every sidereal day only to first order, therefore the elevation and azimuth of a satellite for a specific GPS (sidereal-) day epoch may gradually change over time.
- Seasonal changes take place, which will gradually affect amongst other things, some of the reflectors responsible for multipath (*e.g.* trees).
- Due to variations in solar activity the electron density and distribution of the ionosphere changes in time. This can slightly affect the direction of incidence of GPS signals.

In order to prove or disprove the above statement, a sample autocorrelation analysis was performed. (Autocorrelation is described in Appendix G.) GPS observational data from two IGS stations (HRAO at the Hartebeesthoek Radio Astronomy Observatory and SUTH at the South African Astronomical Observatory, Sutherland) for the whole year of 2002 were processed. The processing residuals for each double-differenced observation from each station to each satellite were extracted and time-stamped according to their sidereal epochs. There were 24879 sets of data

which each contained eighty or more consecutive sidereal days' phase residual data for a specific sidereal epoch and a specific satellite-receiver pair. For each of these sets the sample autocorrelation function (sACF) at lag k was calculated. The average and standard deviation of all the sACFs are given in Table 3.

Table 3: The average sample autocorrelation function at a lag of k days for the whole of 2002, describing the correlation between processing residuals obtained at epochs separated by k sidereal days.

k	sACF
1	0.14 ± 0.15
2	0.11 ± 0.14
3	0.09 ± 0.14
4	0.08 ± 0.14
5	0.08 ± 0.12
6	0.06 ± 0.12
7	0.05 ± 0.11
8	0.03 ± 0.11
9	0.02 ± 0.10
10	0.01 ± 0.16
11	0.02 ± 0.10
12	0.01 ± 0.10
13	0.00 ± 0.10
14	-0.01 ± 0.10
15	-0.01 ± 0.10
16	-0.01 ± 0.09
17	-0.01 ± 0.09
18	-0.02 ± 0.09
19	-0.02 ± 0.09
20	-0.02 ± 0.09

From the low sACF values in Table 3, the conclusion can be made that in most cases the first order effects of multipath and antenna phase centre variations, which are periodic, are too small compared to the other residual contributors (atmospheric variability, GPS noise and higher order effects) to yield a good day-to-day correlation. This is not to say that a good correlation does not exist between the day-to-day effects of multipath and phase centre variations, but only that it might be difficult to detect in the

noisy data. One would, however, still expect unwanted contributors to $\Re$ to be minimised by stacking residuals.

When only considering the data sets for which sACF > 0.5 at $k = 1$ day, *i.e.* the 656 data sets in which the repetitive multipath and antenna phase centre variation effects are clearly visible, the average and standard deviation of the sACFs are given in Table 4.

Table 4: The average sample autocorrelation function at a lag of k days, describing the correlation between processing residuals obtained at epochs separated by k sidereal days. In contrast with Table 3, only data sets for which sACF > 0.5 at k=1 were included in this analysis.

k	sACF
1	0.56 ± 0.05
2	0.47 ± 0.09
3	0.41 ± 0.10
4	0.37 ± 0.11
5	0.32 ± 0.12
6	0.29 ± 0.12
7	0.25 ± 0.14
8	0.20 ± 0.14
9	0.17 ± 0.15
10	0.14 ± 0.15
11	0.11 ± 0.16
12	0.07 ± 0.16
13	0.04 ± 0.17
14	0.01 ± 0.17
15	-0.01 ± 0.17
16	-0.02 ± 0.17
17	-0.03 ± 0.17
18	-0.05 ± 0.16
19	-0.08 ± 0.15
20	-0.09 ± 0.14

The sACF values in Table 4 are graphically presented in Figure 17.

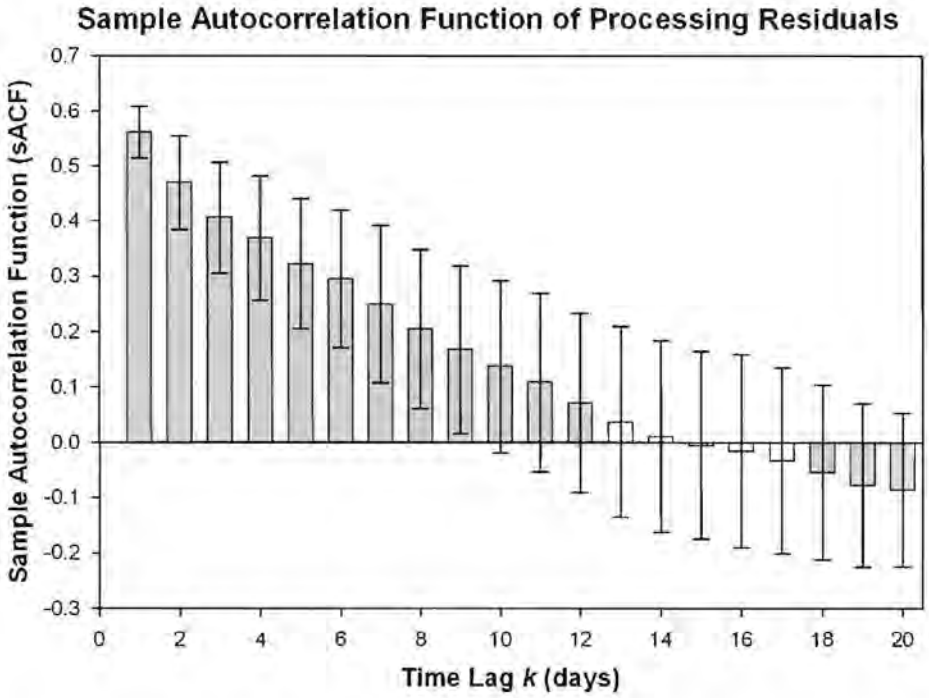

Figure 17: A graphical representation of the average sACF as a function of lag *k*. The presented values are derived from data sets for which there are high day-to-day correlations of the processing residuals.

A modified residual stacking (MRS) method is now suggested, namely a weighted least-squares algorithm, implemented through the weighted multiple regression with dummy variables, as described by equation (F.7). The weights to be included in the analysis will not all be equally set to 1 as in standard residual stacking, but the lag-dependent sample autocorrelations as presented in Table 4 will be used as the weighting function. The number of days prior to and following the day under consideration remains to be selected. Another important difference between this new MRS method and the residual stacking method as used by Braun *et al.* (2001:459–472) is that the day under consideration is not included in the analysis (*i.e.* weighting is set to zero) when the multipath map is generated, so that the multipath map is not contaminated by actual atmospheric variability on the day under consideration.

However, a disadvantage of this proposed method is that it requires the processing residuals at the relevant epochs of the selected number of sidereal days before and after the day under consideration to be available for inclusion in the analysis, while "missing values" are much easier to treat in the case of normal residual stacking.

5.2.4 Comparison between residual stacking and modified residual stacking

GPS data from two IGS stations, namely HRAO and SUTH, for the CONT05 campaign (cf. Section 4.2.2) were processed using the GAMIT suite of software and the same strategy as described in Section 4.2.1 to obtain 2-hourly estimates of ZWD at the two sites. In addition to the estimated zenith delays, processing residuals for each epoch and daily gradients, as defined in (5.2) have also been estimated.

In order to compare the slant delays obtained from GPS and radiometry, the sky above the Hartebeesthoek site was divided into 64 imaginary windows – 8 sectors (azimuth) and 8 rings (elevation). For this study it is assumed that there is a uniform distribution of water vapour in each of these windows, *i.e.* the equivalent zenith delay was assumed to be constant for each window to account for elevation angle dependence. Furthermore, all slant wet delay estimates were mapped to zenith for the purpose of comparison. During the campaign there were 1065 WVR pointings to windows where one or more GPS satellites were being tracked.

To estimate the GPS-derived slant wet delay at each epoch under consideration, the zenith wet delays were linearly interpolated to the epoch under consideration, since they are based on the piece-wise linearly estimated zenith tropospheric delays, and mapped to the relevant elevation angle. In addition to this the daily gradients were applied, based on the azimuth and elevation of the satellite observation. Lastly, the processing residuals $\Re$ were added.

Three approaches to obtain the relevant processing residuals are compared: applying no corrections, applying residual stacking and applying the modified residual stacking (MRS) method. The obtained rms difference between the WVR-derived and GPS-derived slant wet delays (mapped to zenith) when no corrections are applied, is 17.33 mm when considering all 1065 coinciding pointings.

For the application of residual stacking and MRS the number of sidereal days to include in the analysis is set as a variable. In Table 5 the three approaches are compared to each other as a function of the number of days included in the analysis. Recall that a disadvantage of MRS is that processing residuals can only be corrected if the residual data from the selected number of days before and after the day under consideration are available; therefore, the number of pointings compared decreases as the selected number of days increases. Consequently the rms obtained when applying no corrections also changes depending on the number of days included in the analysis, because this determines which pointings will be included in the comparison.

Table 5: A comparison of three different methods to obtain GPS-derived slant wet delays (mapped to zenith).

Number of days in stack	Number of days compared	rms difference (mm) w.r.t. WVR		
		No correction	Residual stacking	MRS
1	14	14.24	13.40	13.60
2	12	13.54	12.78	12.95
3	10	14.06	13.20	13.32
4	8	11.68	11.12	11.10
5	6	11.05	10.41	10.36
6	4	11.36	11.04	10.98
7	2	12.48	12.21	12.14

By applying corrections for multipath and antenna phase centre variation, either through residual stacking or MRS, the rms differences of the GPS-derived slant wet delays (mapped to zenith) with respect to the WVR-derived values were reduced by ~ 0.6

mm. In Figure 18 the results obtained by residual stacking and MRS are compared graphically.

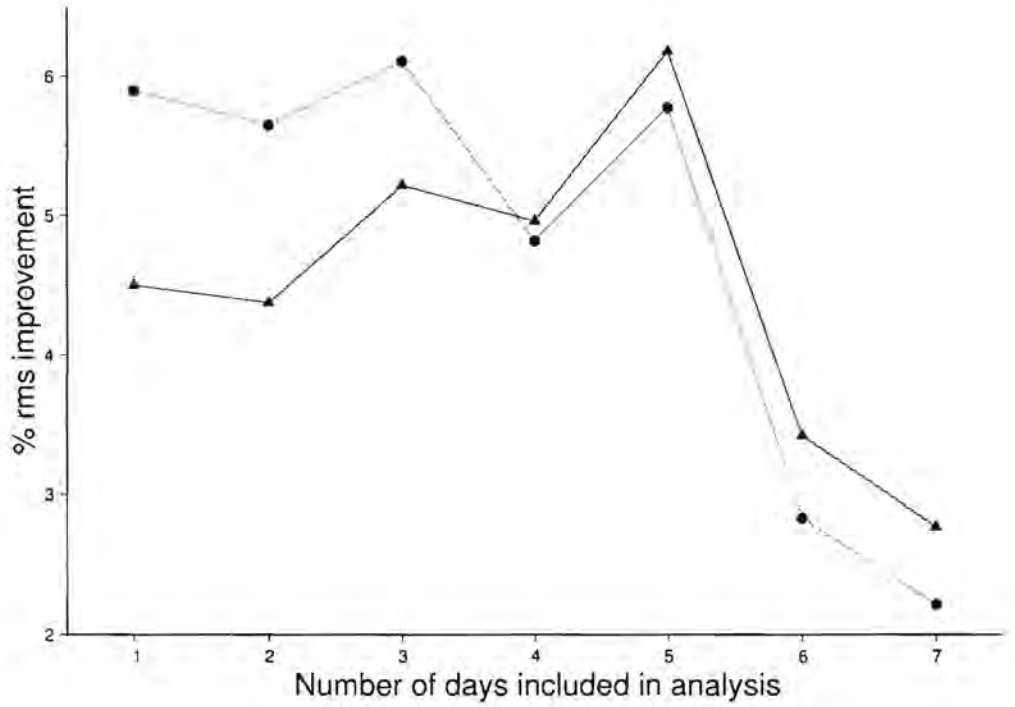

Figure 18: A comparison of the GPS–WVR slant wet delay rms improvements using residual stacking (circles) and modified residual stacking (triangles).

It appears that only in the cases where more than three days of processing residual data before and after the day under consideration were included, did MRS yield better results than residual stacking when compared to radiometry. Since MRS uses a more intelligent weighting of data (cf. Section 5.2.3) it is expected to perform better than residual stacking when more days' data are included.

Only two weeks of continuous GPS and WVR observations may not be sufficient to draw conclusions on whether residual stacking yields better results than MRS. However, a few remarks can be made regarding the comparison presented in Table 5 and Figure 18. Firstly, it does not appear as if MRS provides a significant improvement to GPS-derived slant delays when compared to residual stacking. Secondly, apart from the

disadvantage of being computationally more intensive, MRS has to its disadvantage the fact that much less slant delay data are corrected and produced due to the requirement of processing residuals from days prior and posterior to the day under consideration being available.

At present it does not seem very practical to implement MRS, because the much simpler residual stacking yields comparable results. However, this method may be re-visited in future

- when newer technology will decrease the required computational time;
- with an implementation of MRS which, unlike the multiple regression with dummy variables as described in Appendix F, is well-suited for the handling of missing values;
- when a more accurate source of slant delay data than radiometry becomes available for the purpose of comparison, since the uncertainties of delay measurements made by WVR are ~ 10 mm (A. Somieski, 2006, personal communication); and
- when considering sites more prone to multipath or antennae more prone to phase centre variations than the HRAO site and antenna considered in this study.

6 ANALYSIS OF PWV TIME-SERIES OVER SOUTH AFRICA

6.1 Introduction

Pielke (2005:1626) points out that the effects of environmental change are experienced by people and ecosystems regionally, which necessitates the study of PWV on a regional level (*e.g.* at GPS sites) as opposed to only studying temperature and PWV using global or zonal averages, such as the interpolated grid data available from meteorological analyses (*e.g.* only twenty grid-points over South Africa of the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) Reanalysis Project, which is discussed in Section 6.2).

In this chapter results are presented from a long-term monitoring of PWV over South Africa, using the two South African networks of GPS stations: the Southern Africa Development Community (SADC) GPS Network (which forms part of the International GNSS Service, IGS) and the Trignet Network (which is managed by the Chief Directorate: Surveys and Mapping, South Africa's national mapping agency). Figure 19 shows the distribution of the network analysed in this work, which consists of 30 continuously-operating stations relatively well-distributed over the entire region.

The entire SADC+Trignet network currently processed contains 42 stations. However, a threshold value of ~4 years has been considered in order to have sufficient data to accurately isolate the long-term trend from the other signals present in the daily time-series (*e.g.* seasonal variations). Figure 20 lists graphically the processed data, showing that most of the potential data have been processed.

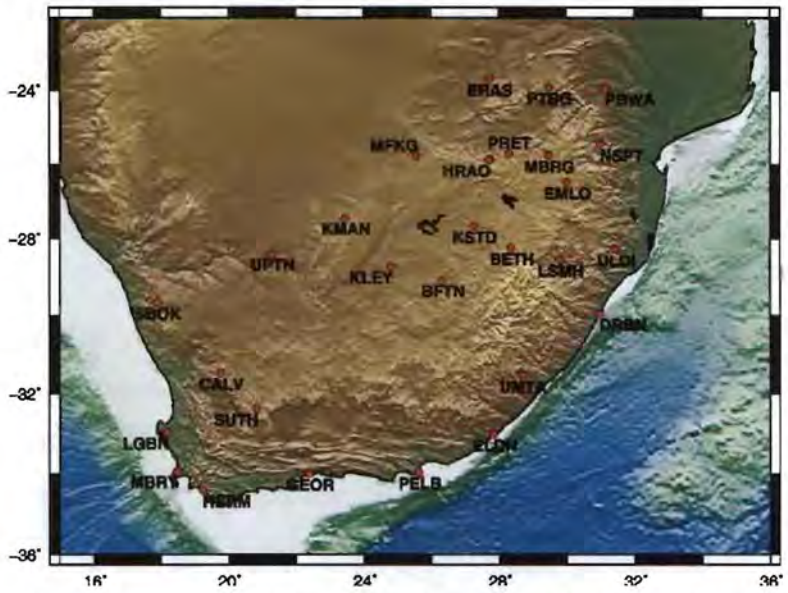

Figure 19: An overview of the 30 continuously-operating GPS stations in South Africa used in this work.

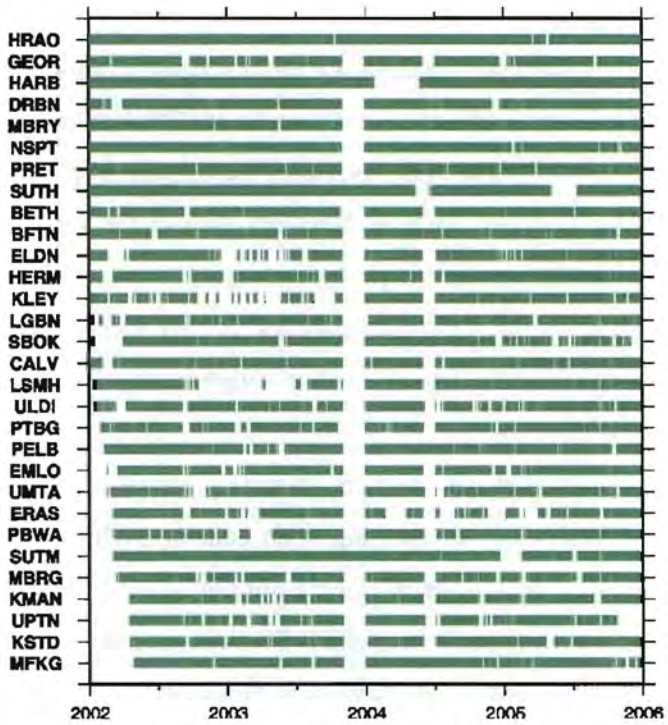

Figure 20: Processed daily files between January 2002 and December 2005 for the GPS stations discussed here. White gaps show periods with no data available, *e.g.* the data loss due to a Trignet server failure during late-2003. Some stations were processed from before 2002 (HRAO processed since October 1996).

The derived PWV trends are compared with simultaneously observed trends in temperature at each of the sites. This analysis allows us to determine to what extent GPS-derived PWV can be utilised to observe temperature change over a relatively short period. (Note that comparisons with precipitation trends were not considered in this study, since precipitation was well below average during 2003 at most stations, biasing precipitation trends for our period of study (2002.0 to 2006.0) to appear strongly positive.) Precipitation occurs sporadically, which makes fitting a trend to the data very dubious because we cannot estimate the uncertainty of the results. Contrary to the temperature and PWV data, it is difficult to stochastically model the precipitation data. The below average precipitation could have an effect on PWV trends; however, since the relationship between PWV and precipitation is complex, this correlation cannot easily be quantified. The explicit aim of this study will be to compare PWV and temperature trends and therefore a further investigation of the effects of precipitation on PWV trends is omitted.

The estimated PWV trend patterns derived from our computations are interpreted in the framework of known features of the climate in South Africa. South Africa's Mean Annual Precipitation (MAP) ranges from less than 100 mm to over 1200 mm, with a mean of ~490 mm. However, approximately one-third of southern Africa has a mean annual runoff of less than 10 mm. Furthermore, when considering the rainfall to runoff conversion, it is found to be less than 5% for approximately half of southern Africa, compared to the world mean of 35% (Schulze, 2005:84). Factors contributing to this low conversion rate include the high levels of aridity, the high ratio between mean annual potential evaporation and MAP, and the high seasonality of rainfall. South Africa is therefore considered to be a "high-risk hydroclimatic environment" (Schulze, 2005:84), making studies of and adaptation to its possible climate changes of utmost importance.

6.2 Methodology

The available GPS data (cf. Figure 20) were processed using the GPS-Inferred Positioning System (GIPSY) software package (Webb & Zumberge, 1995). GIPSY minimises the errors in the distance from satellite to receiver by computing, using a very stable Kalman filter, corrections to the satellite and receiver clocks. This undifferenced approach allows one, when precise orbits and clock corrections are provided and kept fixed, to compute each station position individually, in a strategy known as Precise Point Positioning (PPP) (Zumberge *et al.*, 1997:5005–5017). This strategy has been used to analyse the observations of the South African network since it provides accuracies similar to other approaches based on the simultaneous processing of all observations in networks of equivalent dimensions (Fernandes, 2004). With respect to the parameters of interest, horizontal gradient parameters were estimated to model this effect, since it improves the precision of the estimates in most cases (Bar-Sever *et al.*, 1998:5019–5035, Miyazaki *et al.*, 2003:2335). The Niell (1996:3230) mapping function was used to map tropospheric delays to zenith. Ocean loading (GOT00.2) was also incorporated in the processing (Scherneck & Bos, 2002). For this particular effect, tests have been carried out to quantify its influence in the estimation of the tropospheric parameters. Results have shown that although it can influence the absolute value of the computed zenith tropospheric delay (15 days cycle with a magnitude up to ~10 mm), its effect in the trend estimation is negligible due to the short periodical variation.

The required pressure and temperature estimates at the GPS stations were obtained from the NCEP/NCAR Reanalysis Project (Kalnay *et al.*, 1996:437–471, NCEP, 2006), which aims to produce state-of-the-art global atmospheric analyses based on historical meteorological data.

A few notes should be made about the use of these pressure and temperature data. The surface data, including temperature, pressure and geo-potential altitude, are provided for points on a 2.5×2.5 degree grid. Therefore, data had to be interpolated to the GPS sites of interest.

The strategies used for the vertical interpolation of temperature and pressure estimates to the geo-potential altitude of the GPS sites are based on the environmental temperature lapse rate (6.5K/km) and U.S. Standard Atmosphere 1976 (CRC, 1993:14) respectively. The environmental lapse rate was used because of the short distances over which interpolation was done, the simplicity of implementing this technique and because PWV estimates are not very sensitive to errors in temperature estimates. For interpolation horizontally, the four grid-points closest to the GPS site under consideration were used to calculate an average, which was weighted by the inverse of the distance between the grid-points and the GPS site.

For the purpose of verifying the GPS-derived PWV, PWV grid data were also obtained from the Reanalysis Project. The same technique of horizontal interpolation was applied as in the case of the pressure and temperature estimates, while no vertical interpolation was performed.

The PWV estimates are not very sensitive to errors in temperature estimates: a 10 K error in the temperature estimate will result in ~2.5% error in the scaling of ZWD to PWV. However, PWV estimates are sensitive to pressure estimates; over-estimation of atmospheric pressure will result in under-estimation of PWV and vice versa. Therefore, a comparison between the pressure estimates from the Reanalysis Project and measurements by the South African Weather Service (SAWS) was performed.

The comparison included one year's data (August 2004 to July 2005) for three meteorological stations, namely Irene, George and Cape Town. A mean offset of less than 1 hPa was found between the observed and calculated values at each of the stations. For the two coastal stations, Cape Town and George, the standard deviations of the differences about the mean offset are ~1 hPa, while for Irene the standard deviation is ~0.4 hPa. Therefore, conservative pressure uncertainties of 1.2 hPa and 0.5 hPa for coastal and non-coastal sites respectively were used to calculate the formal uncertainties in PWV estimates. A greater uncertainty is anticipated for coastal stations, since the South African coast is exposed to sudden weather changes due to frontal systems. Note

further that the rms differences between the observed and calculated pressures are not more than ~1.2 hPa in all cases, which corresponds to a total PWV error of ~1.1 mm (Bai & Feng, 2003:85).

At each GPS site one PWV value was estimated for each day, using daily mean estimates of ZTD, pressure and temperature. Following Gradinarsky *et al.* (2002:337) the following four-parameter model was fitted to the PWV data:

$$\mathrm{PWV}(t) = a_1 + a_2 t + a_3 \cos\left[2\pi t + a_4\right] \tag{6.1}$$

with t in years. Therefore an initial offset parameter, a linear trend and the amplitude and phase of a seasonal signal were estimated. A weighted least-squares method, using the squares of the obtained formal PWV uncertainties as the diagonal elements of the covariance matrix, was used for an initial linear regression model fit and a Fourier transformation was performed on the resulting residuals (observations minus model).

It is evident (cf. Figure 21) that the power spectrum of the residuals neatly fits an ARMA(1,1) model (Brockwell & Davis, 1987), defined by

$$X_i = \theta X_{i-1} + \phi Z_{i-1} + Z_i \tag{6.2}$$

where X_i is the observation at time i, Z_i is a Gaussian zero-mean random variable with standard deviation σ, ϕ is the autoregressive (AR) parameter and θ is the moving average (MA) parameter. The parameters for the ARMA(1,1) model depicted in Figure 21 are $\theta = 0.60 \pm 0.02$, $\phi = 0.18 \pm 0.02$ and $Z_i = 3.42$.

The power density P of the ARMA(1,1) model described by (6.2), is given by

$$P(\omega) = \frac{\sigma^2\left(1 + \theta^2 + 2\theta\cos\omega\right)}{2\pi\left(1 + \phi^2 - 2\phi\cos\omega\right)} \tag{6.3}$$

where the frequency $f = \omega/2\pi$. The three parameters describing the ARMA(1,1) model, namely θ, ϕ and σ, are then used to compute a new covariance matrix C using the following algorithm:

$$C_{ij} = \gamma\left(\|i - j\| + 1\right) \tag{6.4}$$

$$\text{with } \gamma(i)=\begin{cases}\sigma^2\left[1+(\theta+\phi)^2/(1-\phi^2)\right] \text{ for } i=1\\ \sigma^2\left[\theta+\phi+(\theta+\phi)^2\phi/(1-\phi^2)\right] \text{ for } i=2\\ \phi\gamma(i-1) \text{ for } i\geq 3\end{cases} \tag{6.5}$$

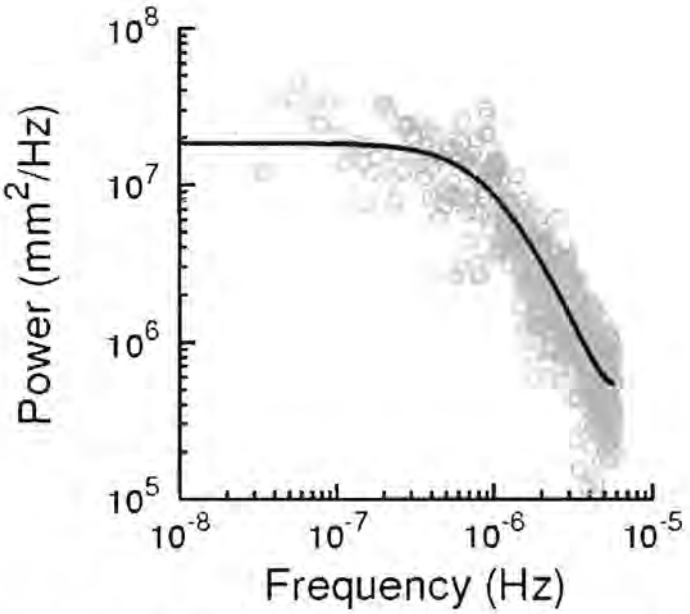

Figure 21: The power spectrum of the residuals of the GPS-derived PWV data for HRAO, after the 4-parameter model was fitted to it using weighted least-squares regression. The solid line depicts the power density for the ARMA(1,1) model, using the obtained parameters for this site.

The new covariance matrix is used in a weighted least-squares analysis in order to determine the four parameters, and realistic uncertainties associated with them, for the model described in (6.1).

The obtained trends are independent of absolute constant biases (Gradinarsky *et al.*, 2002:337) and seasonal influences. The PWV trends can be used to sense long-term changes of PWV.

The temperature data were fitted to the same four-parameter model, using the same strategy as described before. It was also found that the temperature data can be described by an ARMA(1,1) model.

Both the PWV and temperature data are dominated by a large annual signal. Consequently these two parameters are highly correlated. Our objective is to study the changes in the mean PWV and temperature over a period of 4 years; therefore, we do not only fit a linear trend but also an annual signal to the data to separate the long-term changes from the seasonal effect. The use of a proper noise model will ensure that the uncertainties of the estimated trends are realistic. Afterwards the PWV and temperature trends are compared at each station to see if they show any correlation.

Pearson's correlation coefficient r was calculated for the comparison of GPS-derived and Reanalysis-derived PWV. This coefficient was also calculated for the comparison of temperature trends with the PWV-trend at each of the stations.

Lastly, in addition to using the correlation coefficients r to indicate the degree of correlation, use has been made of the probabilities $P_c(r,N)$ that random samples of N uncorrelated experimental data points would yield experimental linear-correlation coefficients as large as or larger than the observed $|r|$, as described in Appendix E.

6.3 Results and discussion

A verification of the GPS-derived PWV by comparison with PWV estimates from the Reanalysis Project is in order before fitting the four-parameter model to the time-series. In the case of all stations considered, correlation coefficients $r > 0.85$ were obtained, thus verifying the described technique.

Figure 22 shows the data, model fit and linear trend of PWV as observed at HartRAO for ~9 years. For the HRAO IGS station located at HartRAO, the obtained

results from the model fit are: linear trend $a_2 = -0.01 \pm 0.07$ mm/yr, seasonal signal amplitude $a_3 = 9.92 \pm 0.25$ mm and seasonal signal phase $a_4 = 5.99 \pm 0.03$ radians.

The model fit was repeated, this time only including data for the 2002.0 – 2006.0 period, during which 30 of the South African GPS stations were observing more or less continuously. The linear trend observed at HartRAO for this period is $a_2 = 0.28 \pm 0.23$ mm/yr.

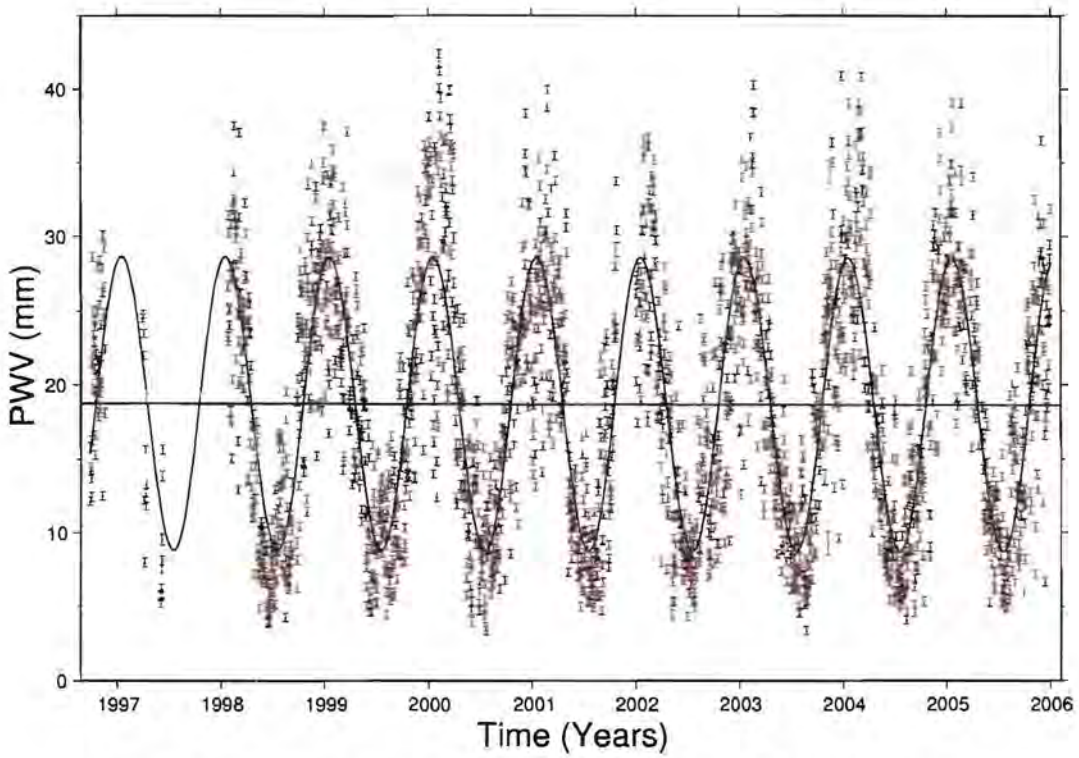

Figure 22: The time-series of GPS-derived PWV for HRAO. The fitted 4-parameter model and the linear trend are also displayed.

Certain global climate cycles with very long periods exist, making it impossible to extrapolate linear trends based on a short period of observations, *e.g.* 4 years, to predict future climate. Also, the trend observed over a specific period, *e.g.* the 9-year period mentioned earlier, would not necessarily represent the trends observed during subsets of that period. For instance, the observed trend for the 4-year period ending December 2001

was 0.27 ± 0.27 mm/yr, compared to the 0.28 ± 0.23 mm/yr for the 4-year period ending December 2005; meanwhile a trend of -0.01 ± 0.08 mm/yr was observed for the 8-year period ending December 2005. The different trends observed over different periods are not indicative of measurement or modelling errors, but of meteorological changes during the observation period. Also, the observed trend does not necessarily have to be linear during the period of observations. However, using quadratic or exponential trends do not improve the fit with the data substantially, which makes the linear trend an acceptable model. Our noise model will ensure that the uncertainties of the estimated trends are realistic and that we can determine if a trend is significantly different from zero or not.

A geographical representation of the PWV trends observed at the different stations is presented in Figure 23. Clear and consistent patterns for different regions are observed: PWV increased over the north-eastern parts and the south-western coast during the period under consideration, while a PWV decrease is observed in a band stretching from the West Coast, through the central interior to the southern part of the East Coast.

The strong correlation observed in the PWV trends at neighbouring stations increases our confidence in the observed regional trends.

It is further observed that the estimated amplitudes (a_3) of the seasonal (cosine) signals in PWV are smaller for sites where lower PWV offset values (a_1) are generally observed, and larger for sites with higher PWV offset values. Furthermore, smaller amplitudes are observed for coastal sites than for non-coastal sites with similar PWV offsets. These observation corresponds to our knowledge of the South African climate, where the observed humidity at non-coastal sites vary strongly between the relatively dry winters and relatively wet summers and the oceans have a stabilising effect on the variations at coastal sites.

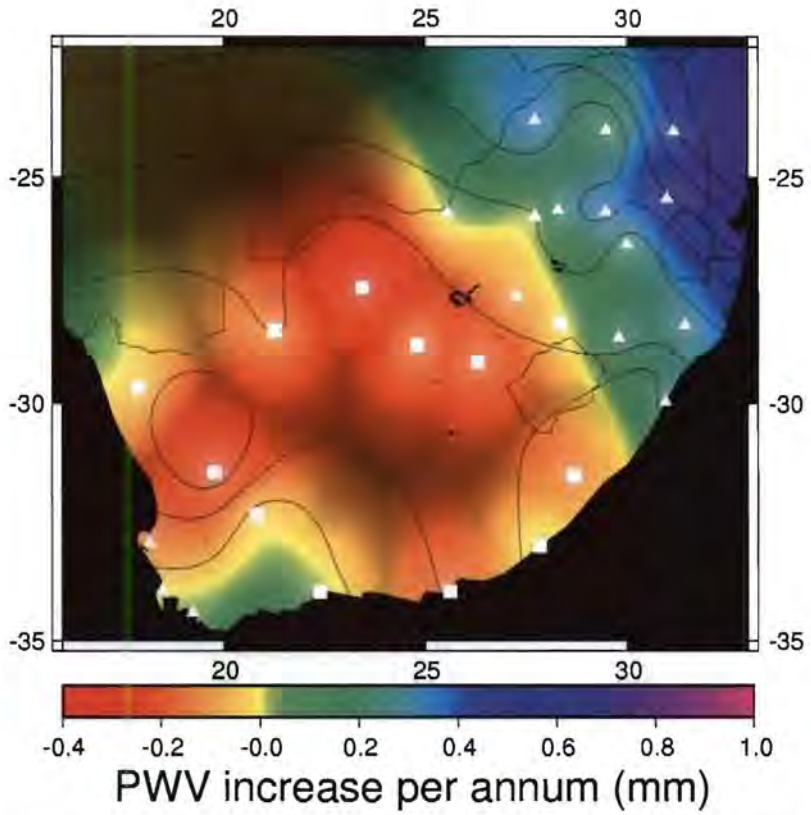

Figure 23: A geographical representation of the observed PWV trends over South Africa for the 4-period of January 2002 to December 2005. Illumination of the interpolated trend values decreases with the distance from the stations. The 1-sigma uncertainties of the estimated trends used to generate this image, have an rms value of 0.21 mm/yr. The triangles represent stations belonging to Group I, which is discussed later, while squares represent Group II stations.

Regarding the phase of the seasonal (cosine) signal, we find $5.29 < a_4 < 6.09$ (in radians) for all stations, which translates to a peak in PWV between 11 January and 26 February, the peak of the austral summer. Higher evaporation rates during this time are due to the higher temperatures, leading to more PWV being observed.

Temperature increases ranging from –0.06 K/yr (at PELB) to 0.28 K/yr (at HRAO) were observed for the 4-year period under consideration. The GPS-derived PWV trends at each of the 30 stations were compared with the trends in temperature for

the corresponding period. All stations were divided into two groups using the following strategy: correlation coefficients (and corresponding probabilities P_c, cf. (E.1)) were calculated for the correlation between temperature residuals and water vapour residuals (*i.e.* observation minus model). In this manner two groups were obtained:

- Group I, with $P_c < 0.32$ – stations for which a strong correlation exists between daily increases (or decreases) in temperature and increases (or decreases) in water vapour.
- Group II, with $P_c > 0.45$ – stations for which no strong correlation exists between daily increases (or decreases) in temperature and increases (or decreases) in water vapour.

None of the stations had values of P_c between 0.32 and 0.45, thereby illustrating the statistical distinctness of the two groups. Correlations were then calculated between the PWV trends and temperature trends for these two groups; the comparisons are presented in Figure 24 and Figure 25.

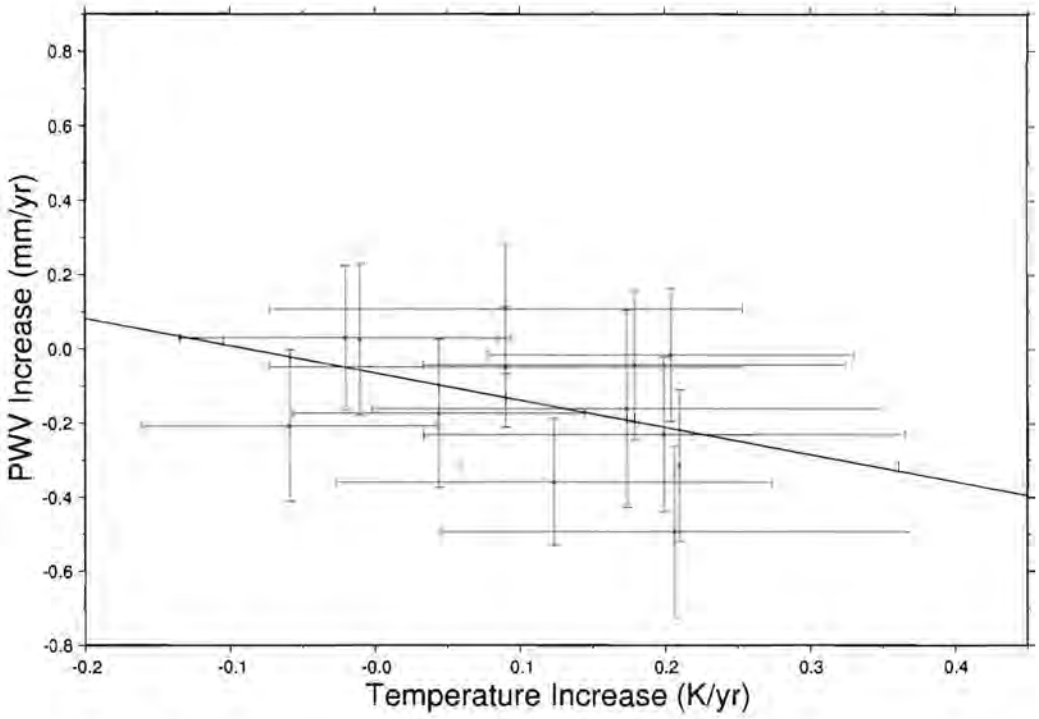

Figure 24: A comparison of PWV and temperature trends for the stations in Group I.

Correlation coefficients $r_I = -0.40$ and $r_{II} = 0.36$ were obtained, yielding probabilities $P_{c,I} = 0.17$ and $P_{c,II} = 0.15$ that temperature trends uncorrelated to the observed PWV trends, would yield the obtained correlation coefficients. The obtained linear regression coefficients were $b_I = -0.73$ mm/K and $b_{II} = 0.71$ mm/K respectively.

It appears that our observations confirm the conclusion by others that only in some cases PWV trends can be used as an indicator of temperature changes, *i.e.* only for some stations temperature increases are correlated to PWV increases. However, the following interpretation provides a possible explanation of our observations, making it possible to use all PWV trends as indicators of temperature changes:

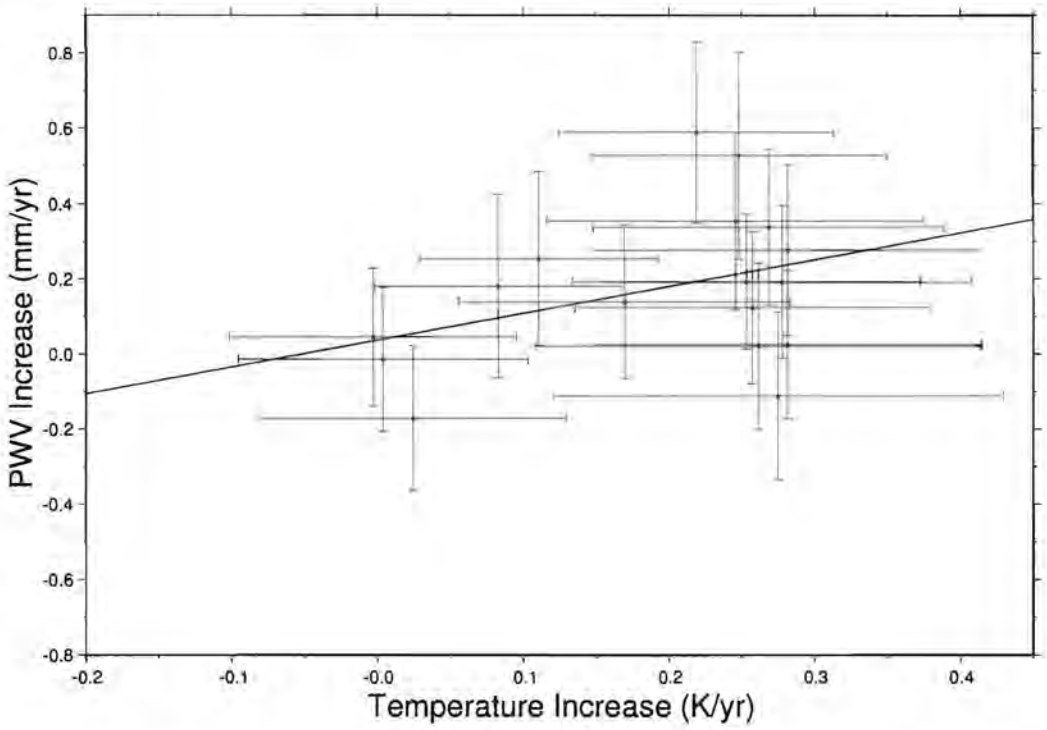

Figure 25: A comparison of PWV and temperature trends for the stations in Group II.

The separation of stations into two groups based on statistical grounds corresponds to what is generally known regarding the South African climate. Stations in Group I are supposedly more dependent on local evaporation for their PWV and consequently a higher probability that temperature residuals and PWV residuals are correlated should be observed for these stations; meanwhile, stations in Group II could depend more on the influx of moisture (water vapour and clouds) from the oceans, the ITCZ (inter-tropic convergence zone) or frontal weather systems. Although a day-to-day increase of temperature at a Group I site might result in a day-to-day increase of PWV, one should ask: what happens when a longer-term (*e.g.* a 4-year period) increase of temperature occurs? The water sources being evaporated might become drier as more and more water is being removed through evaporation, resulting in a gradual decrease of PWV despite an increase in temperature. On the other hand, a longer-term increase in temperature at Group II sites might result in more of the available moisture occurring in the form of water vapour and less liquid water (*e.g.* clouds).

The offsets $9.5 \text{ mm} < a_{1,I} < 25.0 \text{ mm}$ and $16.7 \text{ mm} < a_{1,II} < 30.2 \text{ mm}$ are observed for the two groups respectively, while the mean values of the offsets a_1 are 15.0 mm and 20.8 mm for the non-coastal stations in Groups I and II respectively. (A comparison of the offsets at coastal sites, which are generally higher, was omitted because there were only 3 closely-located coastal sites in Group I.) In general, Group I sites are "drier" while Group II sites are "wetter". Therefore, it might also be possible to divide stations into the distinct groups based on whether or not they can be considered as "dry" or "wet" sites. The Group I sites are mainly concentrated in the more arid central South Africa, where convection is responsible for most of the precipitation, while Group II sites are concentrated in the northeastern and southwestern parts of South Africa (cf. Figure 23), to where the southern branch of the ITCZ and frontal weather systems respectively import large amounts of moisture.

Consequently, PWV trends observed at all stations can be used to investigate the presence of temperature changes, given that one has determined whether the station under consideration has the characteristics of a Group I or II station, *i.e.* whether a negative or

positive correlation exists between the observed PWV trend and the temperature trend at the specific location.

6.4 Summary

It was illustrated how interpolated pressure and temperature data from the NCEP/NCAR Reanalysis Project can be used to obtain precipitable water vapour (PWV) from GPS-derived zenith tropospheric delays (ZTD) within the required accuracies. Thereafter, a previously unanalysed set of GPS data (several years of data from the SADC+Trignet network) was processed to derive time-series of PWV over South Africa.

The derived time-series were fitted to a 4-parameter model and it was illustrated that an ARMA(1,1) noise model, rather than a white-noise model, is required to obtain realistic uncertainties for the obtained parameters.

It was found that PWV over South Africa increased in the northeastern and southwestern parts and decreased in the central parts during our 4-year study period. PWV trends were compared to temperature trends for the same stations during the same study period and it was established that the various GPS sites could be divided into two distinct groups: one for which temperature increase correlated positively with PWV increase and one for which a negative correlation was observed. The first group is located in the arid central part of South Africa. During the 2002.0 – 2006.0 period the temperature increased in this area while PWV decreased. The second group is located in the northeast and southwest of the country where not only the temperature, but also the PWV values increased. The difference between the two groups could be explained within the framework of known features of the South African climate, where the influx of moisture from the oceans to the central plateau is prevented by the mountain ranges separating it from the coastal plains.

7 THE INFLUENCE OF MISMODELLED EARTH TIDES ON GPS-DERIVED PWV

The largest periodic motion of the earth's crust is due to the solid earth tides, which are dominated in the frequency domain by semi-diurnal (M_2, S_2, N_2, K_2) and diurnal (O_1, K_1, P_1, Q_1) constituents (Vey *et al.*, 2002:419). The location-dependent amplitude of the solid earth tide is several centimetres. In Figure 26 the vertical and horizontal components of the earth tide at Sutherland, as per the IERS (2003) model, are presented for a 13-day period. From this figure it is evident that the magnitude of the vertical component of the solid earth tide is almost an order of magnitude greater than for the horizontal components.

It is therefore expected that the solid earth tide needs to be accurately modelled for the purposes of high-precision geodetic applications. Furthermore, if one wishes to use GPS-derived PWV for meteorological applications, accurate models of earth tide are required to obtain the required accuracies for the meteorological applications. Vey *et al.* (2002:425) sets this accuracy requirement at 1 millimetre.

A major limitation of current solid earth tide models is their inability to take into account the spatially varying elastic response of the earth's surface, which depends on the crust's local structure and rheology.

In the following section we will compare the predicted vertical earth tide with super-conducting gravimeter vertical earth tide measurements at Sutherland in order to study the effects of mismodelled vertical earth tides on GPS-derived PWV. Thereafter, a number of different strategies to overcome the effects of mismodelled earth tides will be compared.

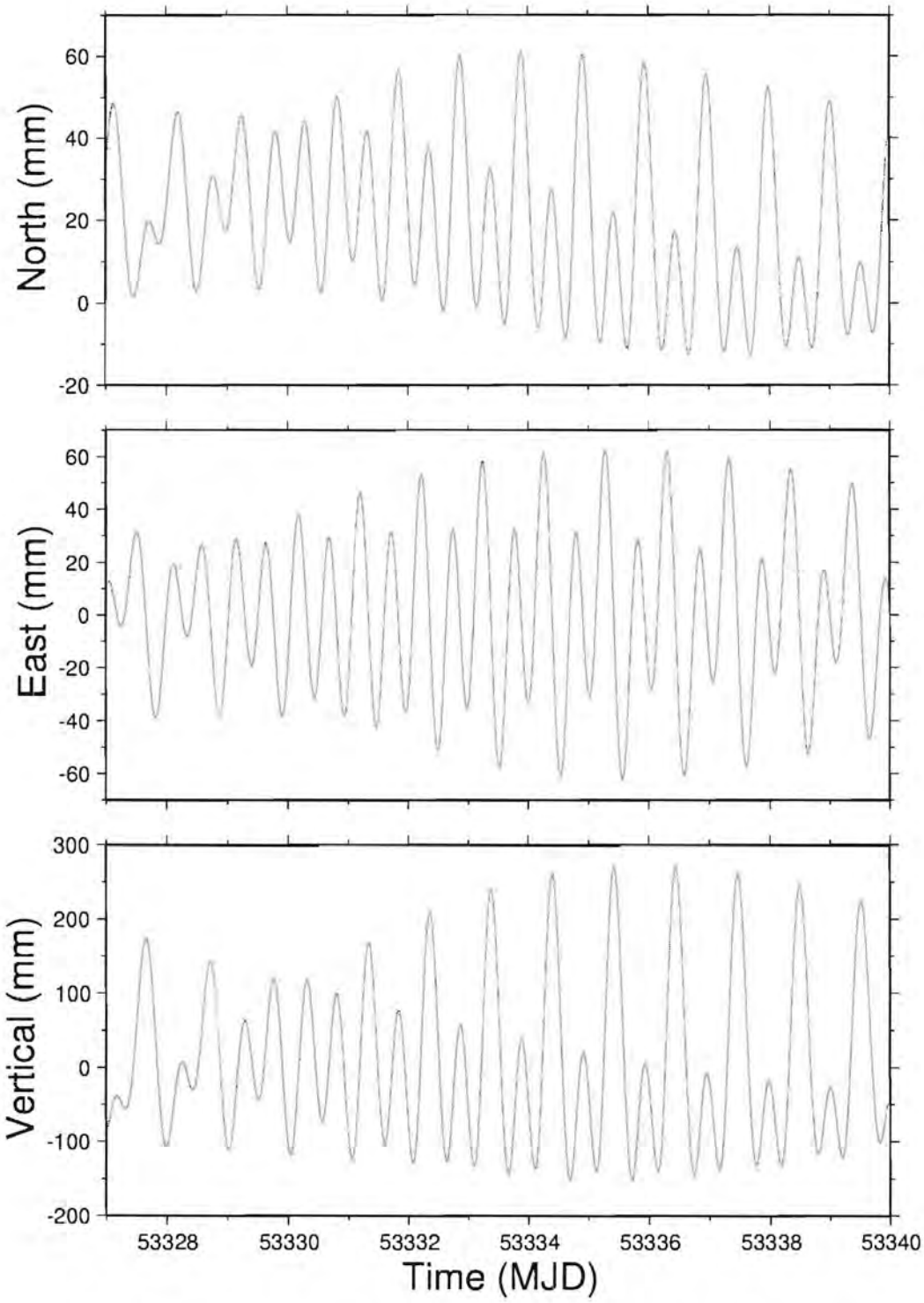

Figure 26: The horizontal and vertical earth tide displacements for Sutherland, as per the IERS (2003) model.

7.1 Comparison with super-conducting gravimeter data

Gravity and atmospheric pressure measurements from the super-conducting gravimeter (SG) at Sutherland, collocated with the IGS GPS station SUTH, were obtained from the Global Geodynamics Project (GGP, http://ggp.gfz-potsdam.de).

Gravity changes Δg induce vertical surface displacements ξc. This shift can be calculated by

$$\xi c = -\Delta g \frac{Rh_2}{2g\delta_2(1+k_2)} \tag{7.1}$$

with the geocentric radius $R = 6373830.451$ m determined by tidal analysis, the Love number for elastic deformation $h_2 = 0.6137$, the Love number for deformation potential $k_2 = 0.3041$, the gravimetric factor $\delta_2 = 1.159$ and the gravitational acceleration $g = 9.79079$ m.s^{-2} determined by absolute gravity measurements (Neumeyer *et al.*, 2006). From (7.1) the elastic deformation coefficient for the Sutherland SG has been determined to be

$$\Delta \text{vs} = -\frac{Rh_2}{2g\delta_2(1+k_2)} = -1.32 \text{ mm}/\mu\text{gal} \tag{7.2}$$

In order to calculate the vertical shift, the gravity changes Δg have to be corrected for atmospheric pressure before being multiplied by Δvs. Neumeyer *et al.* (2006) have calculated the atmospheric pressure admittance coefficient apc = –2.92 μgal/hPa for the SG at Sutherland.

Figure 27 presents the vertical earth tide as observed at Sutherland by gravimetry for the 13-day period of 18 to 30 September 2004 and it is compared to the IERS (2003) model in Figure 28.

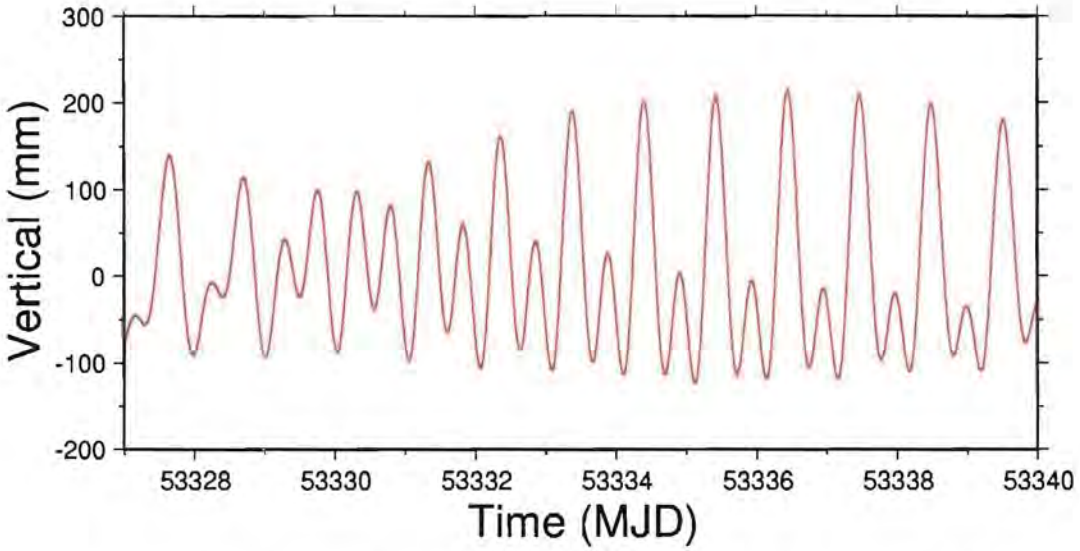

Figure 27: The vertical earth tide displacement observed by gravimetry at Sutherland, corrected for atmospheric loading. The same axis-scaling is used as for the modelled vertical displacements in Figure 26.

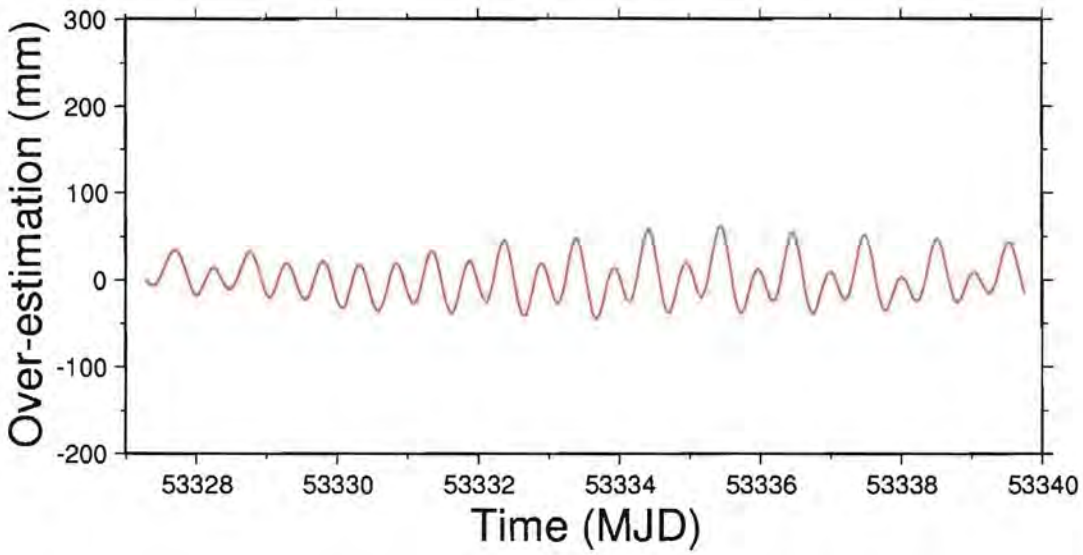

Figure 28: The over-estimation (*i.e.* model minus gravimetry-observed) of vertical earth tide displacements at Sutherland for our period of study.

GPS data for a 13-day period (18 to 30 September 2004) from 6 continuously operating stations were processed using GAMIT (version 10.2). The network of stations is shown in Figure 29. Precise coordinates in the 2000 International Terrestrial Reference

Frame (ITRF2000) for these stations were obtained from R.M.S. Fernandes (2006, personal communication).

The solution estimated tightly constrained station coordinates, satellite state vectors, 25 tropospheric zenith delay parameters per day per site, and phase ambiguities. A ten-degree elevation angle cut-off and IGS final orbits were used. Furthermore, elevation-dependent antenna phase centre variations and IERS (2003) models for implementing solid earth and polar tides were applied. The ZTD variation was constrained to be a Gauss-Markov process and thus care had to be taken to avoid border effects (Vey *et al.*, 2002:423); therefore, only ZTD data from the middle 12-hour period of every 24-hour session were extracted before moving the window forward by 12 hours.

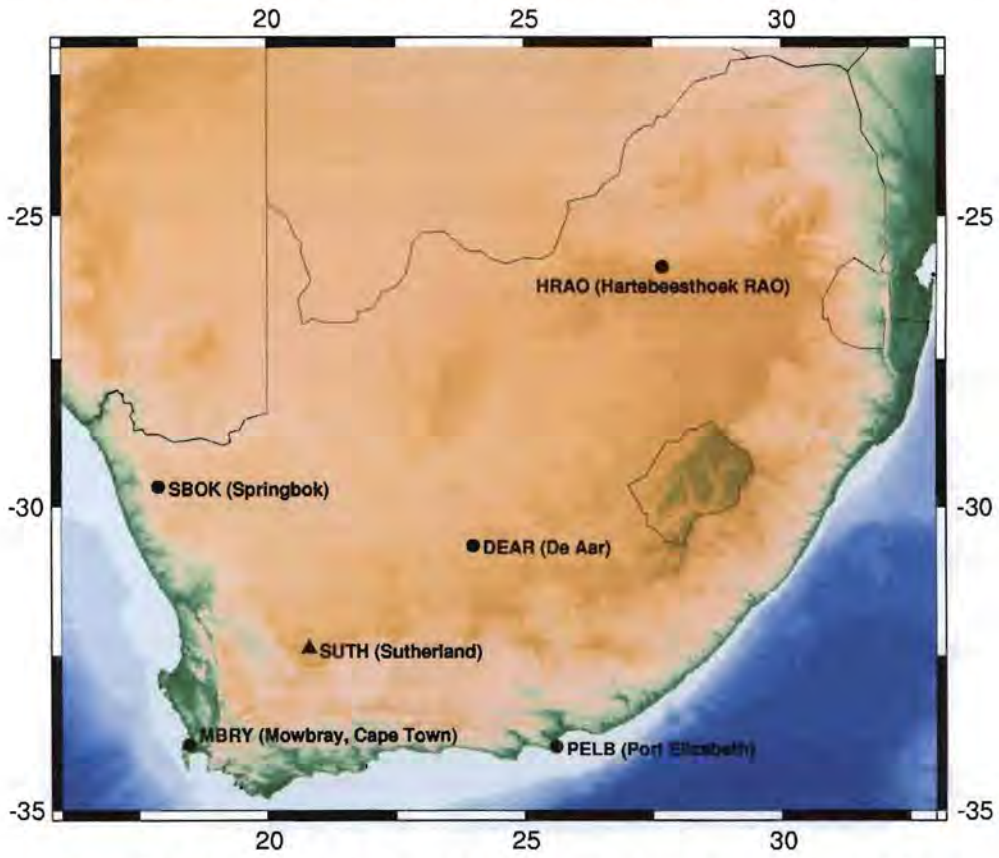

Figure 29: The network of six GPS stations used during this study. SUTH is collocated with a super-conducting gravimeter.

The whole processing procedure was then repeated, but this time using the measured vertical displacement for SUTH, as from the SG data, for each 60-second epoch instead of calculating it from the earth tide model. In both cases a 60-second sampling interval was used instead of the conventional 30-second sampling, because the provided SG data had a 1-minute data rate and additional errors would be introduced by interpolation.

Figure 30 presents the difference in estimated ZTD between the two processing strategies. The obtained correlation and regression coefficients for the applied linear regression are $r = 0.99$ and $b = -0.22$ respectively. The inverse regression coefficient of $b' = 4.4$ corresponds exactly to the value of 4.4 obtained by Vey *et al.* (2002:424), also using a ten-degree elevation angle cut-off, in their study of ocean loading impacts on atmospheric delay estimations. Recalling that the ratio between ZTD and PWV is approximately 6:1, this implies that a 26 mm error in the used earth tide model will introduce a 1 mm error in the obtained PWV. During our period of study we find maximum and rms values of model errors of 62.1 mm and 23.7 mm respectively. These translate into significant errors when considering the required 1 mm PWV accuracy for meteorological purposes (Vey *et al.*, 2002:425).

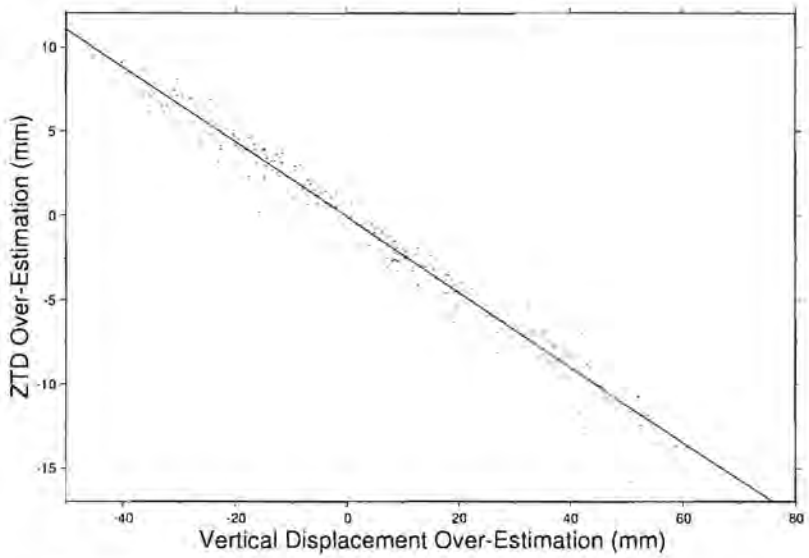

Figure 30: ZTD errors are strongly anti-correlated to the vertical position errors introduced by mismodelled earth tides.

Therefore, in the following section we consider a number of strategies to improve the impact of mismodelled vertical earth tide on GPS-derived ZTD estimates.

7.2 Strategies to reduce the influence of mismodelled vertical earth tides on GPS-derived ZTD

From Figure 31 it is evident that scaled predicted vertical earth tide displacements (VETD) would better resemble the observed VETD at Sutherland, *e.g.* using a linear regression coefficient (in the comparison of observed VETD with modelled VETD) as the scaling parameter.

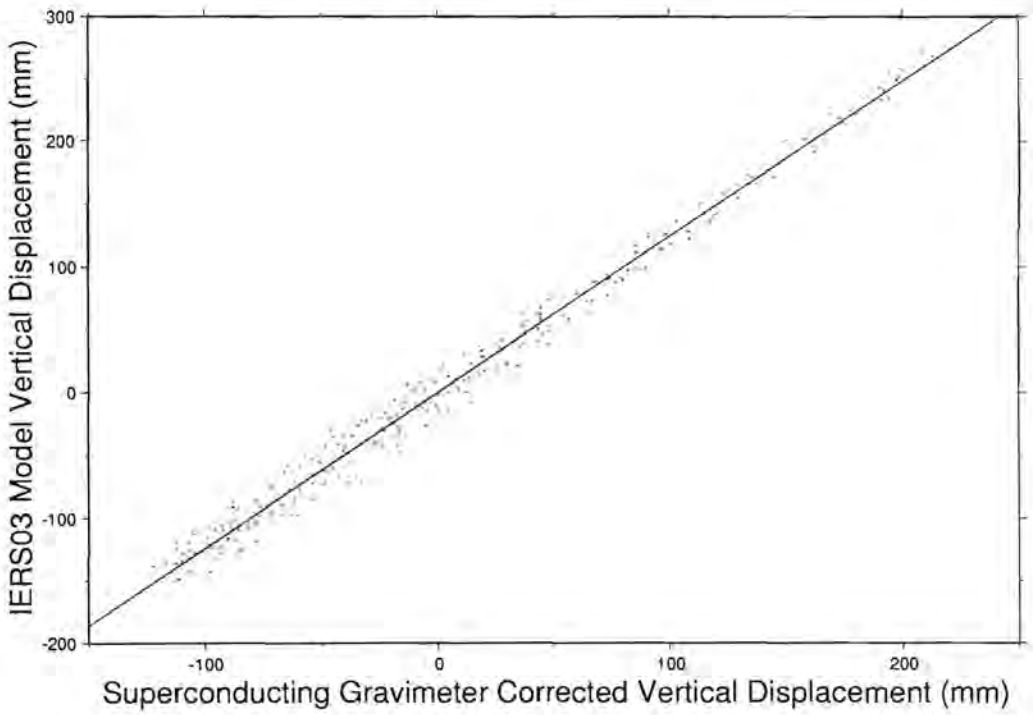

Figure 31: A comparison of modelled (by IERS 2003 model) and measured (by gravimetry) vertical earth tide displacements. Note that only hourly data are presented in this graph, while 1-minute interval data were used to calculate the regression coefficient.

A scaling factor (regression coefficient) $b = 0.798$ is obtained for the whole period under consideration. However, this scaling parameter would not necessarily be fixed for the whole period; Table 6 presents the scaling factors obtained for the individual days during our study.

Table 6: A summary of the daily scaling factors between the modelled and the observed vertical earth tide displacements at Sutherland.

MJD	*b*
53327	0.843
53328	0.825
53329	0.801
53330	0.761
53331	0.763
53332	0.767
53333	0.780
53334	0.785
53335	0.790
53336	0.801
53337	0.808
53338	0.818
53339	0.824

In this section we investigate the following strategies to determine the appropriate scaling parameter and investigate its impact on the consequently estimated ZTD values:

- SG-A – using the superconducting gravimeter data for all the days in our study, a single scaling parameter for the whole period is obtained and applied to the modelled VETD. The obtained scaling factor is 0.798 as stated previously.
- SG-1 – using the superconducting gravimeter data for one day at a time to obtain a scaling parameter for that day. The obtained scaling parameters are presented in Table 6.

- GN-A – using the GPS network to measure Sutherland's VETD for all the days in our study to obtain a single scaling parameter for the whole period and apply it to the modelled VETD.
- GN-1 – using the GPS network to measure Sutherland's VETD for one day at a time to obtain a scaling parameter for that day.

The benchmark to be used for comparisons in all cases will be the ZTD obtained using the gravimetry-observed VETD for vertical coordinate corrections at each epoch during processing, as described in Section 7.1.

To obtain the scaling factors for strategies GN-A and GN-1, the same processing strategy as described in Section 7.1 was used to measure the vertical displacement at Sutherland, using the GPS network presented in Figure 29, with the following changes:

- The tightly constrained station parameters were kept fixed, with the exception of unconstraining the vertical component of SUTH's coordinates.
- The 24-hour processing window was reduced to a 5-hour window with a 30-minute step between processing sessions.
- The IERS (2003) earth tides were applied for both the horizontal and vertical components at all stations, with the exception of SUTH's vertical component.
- A 30-second sampling rate was used.
- Three ZTD parameters were estimated per processing session per station.

From each processing session an estimated vertical displacement for SUTH was obtained.

This processing strategy is based on the comparison of VETD measured by GPS and gravimetry by Neumeyer *et al.* (2002:10607–10616). Significant changes to their strategy for this current study include:

- The application of earth tide modelling for the horizontal components at all sites and the vertical components at all sites except the site of interest (*i.e.* SUTH).
- The use of a good network geometry.

The modelled (IERS 2003) VETD and the VETD measured by GPS are compared to the VETD from gravimetry in Figure 32. The rms difference between the VETD measured by gravimetry and GPS is 21.692 mm, which is slightly better than for the IERS (2003) model. It seems that only during some periods, *e.g.* modified Julian days (MJD) 53331 to 53337, the earth tide measured by GPS is more accurate than the model.

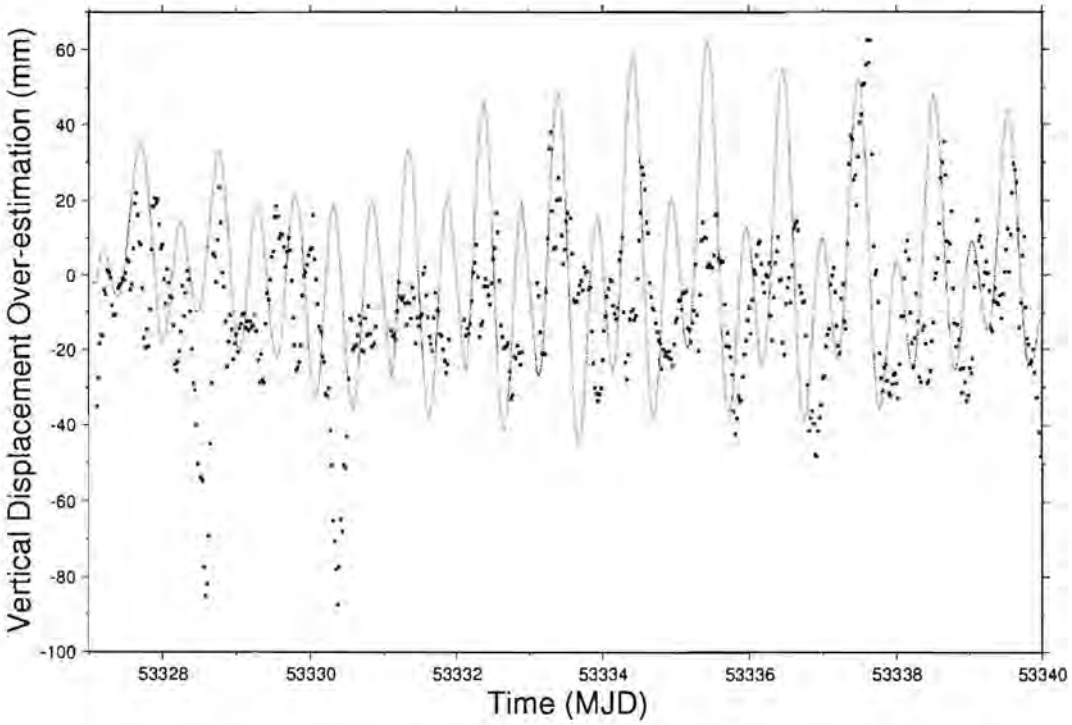

Figure 32: A comparison of the modelled (IERS 2003) and measured (GPS) VETD with the VETD measured by gravimetry. The solid line represents the modelled displacement over-estimation while the dots represent the displacements from the GPS analysis.

A scaling factor (regression coefficient) $b = 0.829$ is obtained for the whole period under consideration, using the GPS-derived VETD estimates. Table 7 presents the scaling factors obtained by strategy GN-1. A comparison of the ZTD obtained by these four strategies, as well as the strategies using no earth tide model and the IERS (2003) model, is summarised in Table 8.

Table 7: A summary of the daily scaling factors between the modelled and the GPS measured vertical earth tide displacements at Sutherland.

MJD	*b*
53327	0.875
53328	0.804
53329	0.768
53330	0.556
53331	0.738
53332	0.775
53333	0.840
53334	0.780
53335	0.851
53336	0.814
53337	0.966
53338	0.826
53339	0.887

Table 8: A comparison of the ZTD estimates obtained using six different strategies. The rms and offset values are calculated relative to the ZTD obtained by using the vertical earth tide displacements as measured by gravimetry, while the improvements are given with respect to the ZTD rms obtained using the IERS (2003) model.

Strategy	Maximum error (mm)	Worst day (MJD)	rms (mm) on worst day	Mean ZTD offset (mm)	rms (mm)	% Improvement
IERS 2003	10.4	53337	6.75	-0.47	5.42	0.0%
No model	55.6	53335	24.96	0.49	20.26	-273.9%
SG-A	4.8	53332	2.25	-0.11	1.86	65.7%
SG-1	4.7	53330	2.09	-0.07	1.77	67.3%
GN-A	4.8	53332	2.61	-0.18	2.02	62.8%
GN-1	6.9	53337	5.82	-0.20	2.69	50.4%

7.3 Discussion of results

From Table 8 it appears that ZTD estimated during GPS processing using a global earth tide model, such as IERS (2003), will result in errors of significant magnitude when considering the consequently derived PWV for inclusion in numerical weather models. Recall that a 5.42 mm rms error in ZTD would result in ~0.9 mm error in PWV; the consequent total PWV error, taking other errors (*e.g.* pressure measurement errors) into account, would be in excess of 1 mm, which is the accuracy requirement for meteorological applications.

The ZTD rms errors in Table 8 refer to the errors at individual epochs of ZTD estimation, generally done at hourly or 2-hourly intervals. When considering the mean daily ZTD for the period under consideration, we find a ZTD rms of 1.03 mm (using the IERS 2003 model) when compared to the daily mean ZTD derived using vertical displacements from gravimetry. The corresponding mean ZTD offset for the period was –0.52 mm. This leads to the conclusion that daily mean ZTD values estimated using the IERS (2003) model are of sufficient accuracy to include in PWV-timeseries studies, such as the study presented in Chapter 6.

A number of strategies for the modelling of earth tides, with the purpose of estimating ZTD, have been presented and compared. Since it would not be feasible (for practical reasons) to operate a gravimeter at every GPS site worldwide, strategies GN-A and GN-1 provide a novel technique to estimate vertical earth tide displacements.

A current shortcoming of these two strategies is that a good GPS network geometry is required, *i.e.* GPS stations should be well-distributed around the site of interest. Furthermore, it should be assumed that the modelled vertical earth tide displacements at the other GPS stations in the analysis are relatively accurate. These limitations are due to the fact that differential (GAMIT) processing was performed to estimate vertical positions. Future research could include an investigation into the use of

precise point-positioning, rather than double-differencing, to measure vertical earth tide displacements.

A number of future improvements could be made to the presented comparison between GPS-derived and modelled VETD:

- Neumeyer *et al.* (2004:449) point out that the pressure data used in the analysis of gravimeter data presented here, are sufficient to model crustal deformation, but that three-dimensional pressure data are required to model the so-called "attraction" term in the gravimeter data. Although this term is an order of magnitude smaller than the deformation term, it still needs to be taken into account when applying corrections to high-accuracy gravimeter data.
- Additional tide gauge measurements could improve the ocean loading corrections applied to gravimeter data (Neumeyer *et al.*, 2005:110). The establishment of a scientific tide gauge network on the southern African coast is currently being considered and could contribute to the accuracy of corrected gravimeter data.
- Only scaling factors were applied to the modelled VETD during the presented analysis because of the short time-series and the comparatively low temporal resolution of GPS-derived VETD data. One might also be able to include a phase lag (or phase advance) term in the model, derived from longer time-series.

The purpose of this study is not to criticise or discredit global earth tide models, in particular the IERS (2003) model, which was considered. This study simply points out that no global model is able to take local rheology into account perfectly and that localised studies might be able to improve earth tide modelling.

8 CONCLUSIONS

The main focus of the work presented in this thesis is the capability of the Global Positioning System (GPS) to sense atmospheric precipitable water vapour (PWV).

The presented literature review includes an overview of GPS and its basic concepts and components, as well as the theoretical background to some of the methods used to observe tropospheric water vapour and ionospheric electron content. For the more theoretical-minded reader derivations of the tropospheric and ionospheric delays experienced by electromagnetic waves traversing the Earth's atmosphere from first physical principles, were presented in the appendices.

Four of the PWV-related research foci of the Hartebeesthoek Radio Astronomy Observatory (HartRAO) Space Geodesy Programme were considered and discussed in this thesis.

Firstly, GPS-derived PWV and zenith wet delay (ZWD) estimates were compared to the PWV and ZWD estimated by other techniques, namely very long baseline interferometry (VLBI), water vapour radiometry (WVR) and radiosondes (RS). Radiosondes are launched daily from stations of the South African Weather Service (SAWS), while HartRAO operates the other three techniques. In all cases a high degree of inter-technique correlation was observed, which verifies the validity of the presented GPS-derived PWV estimates and presents the first WVR- and VLBI-derived meteorological products from HartRAO's Space Geodesy Programme.

Secondly, the geodetic use and possible accuracy improvements of slant delay estimates were investigated. Using GPS-derived slant delays at HartRAO and Fundamentalstation Wettzell, Germany in the VLBI analysis proved to improve VLBI inter-station baseline repeatabilities slightly; however, the greater formal errors of the GPS-derived delays yielded slightly greater formal baseline errors. Furthermore, it was

illustrated that by stacking GPS processing residuals one could slightly reduce the rms difference between WVR- and GPS-derived slant delays by making corrections for the effects of multipath and antenna phase centre variations.

A modified residual stacking (MRS) method, which intuitively is more sensible and for which the weighting function was based on measured day-to-day processing residual correlations, was also presented. As expected, MRS yields slightly better results when a larger number of days are included in the analysis, but the requirements for implementing MRS and the insignificant improvements yielded do not make it worthwhile to implement.

Thirdly, the usefulness of GPS-derived PWV for studies of South African climate phenomena was investigated. A four-parameter model, which is widely used to detect linear trends and correct for seasonal effects, was fitted to PWV time-series data of 30 South African stations for the 4-year study period.

An indication of trend reliability, such as a formal uncertainty, is required to interpret the significance of the obtained trends. It was illustrated that the power spectra of both the PWV and temperature residuals fit an autoregressive moving average (ARMA) model rather than a white-noise model. It has therefore been established that an ARMA model is more suitable for the estimation of trend uncertainties of this kind than a white-noise model, which is implicit in a normal weighted least-squares analysis.

For the period under consideration, January 2002 to December 2005, significant non-zero PWV trends have been detected over certain regions of South Africa. Furthermore, the detected PWV trends showed a spatial correlation, implying an increase in PWV over the northeast and southwest of the country and a decrease over the central areas. This was observed during a period when the whole country experienced an increase in temperature.

Dividing South African GPS stations into two groups, based on the local sources of PWV, the relationship between PWV and temperature trends could be explained. A hypothesis was presented, stating that PWV trends would be negatively correlated with temperature trends in areas where local evaporation was the main source of PWV, while a positive correlation could be expected in areas where the PWV mainly has external sources.

Fourthly, the effects of mismodelled vertical earth tide displacements (VETD) on zenith delay estimates were considered. In particular, the super-conducting gravimeter measurements at Sutherland, South Africa were compared to the IERS (2003) model and it was shown that mismodelled VETD could contribute to significant errors in PWV estimates.

A method was presented to measure the vertical displacement of stations due to the earth tide, using GPS. In some cases this proved to be more accurate, compared to collocated gravimeter measurements, than the applied global model. Apart from its possible use in geodesy, it was shown that this technique could be used to significantly improve the estimates of PWV in areas where the vertical earth tide is not well modelled.

It is evident that GPS could be useful to radio astronomers, geodesists, meteorologists and climatologists as a sensor of atmospheric water vapour. Due to the financial implications it does not seem likely that a very dense network of GPS receivers, dense enough to perform tropospheric tomography, will soon be established in South Africa. However, the ongoing augmentation of GPS occultation satellite constellations, such as the six COSMIC (Constellation Observing System for Meteorology, Ionosphere and Climate) satellites launched in April 2006, will continue to increase the amount of data available to retrieve vertical profiles of atmospheric water vapour. The combination of occultation-derived profiles and ground-based measurements of integrated water vapour could, in future, result in an accurate near real-time three-dimensional PWV product.

REFERENCES

ALBER, C., WARE, R., ROCKEN, C., BRAUN, J. 2000. Obtaining single path phase delays from GPS double differences. *Geophysical Research Letters.* **27**(17):2661–2664.

ASHBY, N. 2002. Relativity and the Global Positioning System. *Physics Today.* May 2002:41–47.

BOEHM, J., SCHUH, H. 2004. Vienna Mapping Functions in VLBI analyses. *Geophysical Research Letters.* **31**(17):L01603.

BAI, Z., FENG, Y. 2003. GPS Water Vapor Estimation Using Interpolated Surface Meteorological Data from Australian Weather Stations. *Journal of Global Positioning System.* **2**(2):83–89.

BAR-SEVER, Y.E., KROGER, P.M., BORJESSON, J.A. 1998. Estimating horizontal gradients of tropospheric path delay with a single GPS receiver. *Journal of Geophysical Research.* **103**:5019–5035.

BEVINGTON, P.R. 1969. Data Reduction and Error Analysis for the Physical Sciences. New York: McGraw-Hill. 336pp.

BEVIS, M., BUSINGER, S., HERRING, T.A., ROCKEN, C., ANTHES, R.A., WARE, R.H. 1992. GPS Meteorology: Remote Sensing of Atmospheric Water Vapour Using the Global Positioning System. *Journal of Geophysical Research.* **97**(D14):15787–15801.

BORBÁS, É. 1997. Determination of Precipitable Water for a fixed site using the Global Positioning System technique. *Quarterly Journal of the Hungarian Meteorological Service.* **101**(4):261–273.

BOSLOPER, C. 1990. Multipath and GPS. University of New South Wales: Technical Report. 206pp.

BRAASCH, M.S. 1995. Multipath Effects. *Global Positioning System: Theory and Applications.* **I**:547–568. Eds. Parkinson, B.W., Spilker, J.J. Jr. Reston: AIAA.

BRAUN, J., WARE, R., ROCKEN, C. 2001. Validation of line-of-sight water vapour measurements with GPS. *Radio Science.* **36**(3):459–472.

BROCKWELL, P.J., DAVIS, R.A. 1987. Time series: theory and methods (2nd edition). New York: Springer.

CHEN, F.F. 1984. Introduction to Plasma Physics and Controlled Fusion (2nd edition). New York: Plenum Press. 421pp.

CHOUDHURI, A.R. 1998. The Physics of Fluids and Plasmas. Cambridge: University Press. 427pp.

CILLIERS, P.J., GOUWS, D., OPPERMAN, B., WONNACOTT, R.T., COMBRINCK, L. 2003. The South African network of dual-frequency global positioning system satellite receiver base stations: a national asset with many applications and research opportunities. *South African Journal of Science.* **99**:51–55.

COMBRINK, A.Z.A. 2003. Detection of atmospheric water vapour using the Global Positioning System. Potchefstroom University for CHE: M.Sc. Dissertation. 60pp.

COMBRINK, A.Z.A., COMBRINCK, W.L., MORAAL, H. 2004. Near real-time detection of atmospheric water vapour using the SADC GPS network. *South African Journal of Science.* **100**:436–442.

CRC. 1993. U.S. Standard Atmosphere (1976). *CRC Handbook of chemistry and physics: a ready-reference book.* **14**:12–19. Ed. Lide, D.R. Boca Raton: CRC Press.

DAVIS, J.L., HERRING, T.A., SHAPIRO, I.I., ROGERS, A.E.E., ELGERED, G. 1985. Geodesy by radio interferometry: Effects of atmospheric modelling errors on the estimates of the baseline length. *Radio Science.* **20**(6):1593–1609.

DEBYE, P. 1929. Polar Molecules. New York: Dover. 200pp.

EOM (EARTH OBSERVATION MAGAZINE). 1998. GPS Q&A: Multipath. *Earth Observation Magazine.* **7(10)**. [Internet:] http://www.eomonline.com/Common/Archives/1998oct/98oct_gqa.html. [Date of access: 4 May 2005]

EMARDSON, T.R. 1998. Studies of atmospheric water vapour using the Global Positioning System. Chalmers University of Technology: Technical Report. 29pp.

FERNANDES, R.M.S. 2004. *Present-day kinematics at the Azores Gibraltar plate boundary as derived from GPS observations.* Delft: DUP. ISBN 90-407-2557-8

FLORES, A., DE ARELLANO, J.V.-G., GRADINARSKY, L.P., RIUS, A. 2001. Tomography of the Lower Troposphere Using a Small Dense Network of GPS Receivers. *IEEE Transactions on Geoscience and Remote Sensing.* **39**:439–447.

GRADINARSKY, L.P. 2000. Remote sensing of small-scale structures of atmospheric water vapor. Chalmers University of Technology: Technical Report. 56pp.

GRADINARSKY, L.P. 2002. Sensing atmospheric water vapor using radio waves. Chalmers University of Technology: Ph.D. Thesis. 150pp.

GRADINARSKY, L.P., JOHANSSON, J.M., BOUMA, H.R., SCHERNECK, H.-G., ELGERED, G. 2002. Climate monitoring using GPS. *Physics and Chemistry of the Earth.* **27**:335–340.

GRIFFITHS, D.J. 1999. Introduction to Electrodynamics (3rd edition). Upper Saddle River: Prentice Hall. 576pp.

HOFFMANN-WELLENHOF, B., LICHTENEGGER, H., COLLINS, J. 1993. GPS – Theory and Practice (2nd edition). New York: Springer Verlag. 326pp.

IERS (INTERNATIONAL EARTH ROTATION AND REFERENCE SYSTEMS SERVICE). 2003. IERS Conventions 2003. Eds. McCarthy, D.D., Petit, G. Frankfurt am Main: BKG.

IGS (INTERNATIONAL GNSS SERVICE). 2005. Data & Products. [Internet:] http://igscb.jpl.nasa.gov/components/prods.html. [Date of access: 21 January 2005]

JANES, H.W., LANGLEY, R.B., NEWBY, S.P. 1991. Analysis of Tropospheric Delay Prediction Models: comparisons with ray-tracing and implications for GPS relative positioning. *Bulletin Géodésiqué.* **65**(**3**):151–161.

JEKELI, C. 2001. Inertial Navigation Systems with Geodetic Applications. Berlin: Walter de Gruyter. 352 pp.

KALNAY, E., KANAMITSU, M., KISTLER, R., COLLINS, W., DEAVEN, D., GANDIN, L., IREDELL, M., SAHA, S., WHITE, G., WOOLLEN, J., ZHU, Y., CHELLIAH, M., EBISUZAKI, W., HIGGINS, W., JANOWIAK, J., MO, K.C., ROPELEWSKI, C., WANG, J., LEETMAA, A., REYNOLDS, R., JENNE, R., JOSEPH, D. 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**:437–471.

KAPLAN, E.D. 1996. Understanding GPS: Principles and Applications. Boston: Artech House. 554pp.

KING, R. 1997. Documentation for the GAMIT GPS Analysis Software – Atmospheric Delay Models. [Internet:] ftp://bowie.mit.edu/pub/gps/updates/documentation/PDF_files/chap8.pdf. [Date of access: 10 January 2003]

KING, R. 2000. Documentation for the GAMIT GPS Analysis Software – Overview of GAMIT Processing. [Internet:] ftp://bowie.mit.edu/pub/gps/updates/documentation/PDF_files/chap1.pdf. [Date of access: 10 January 2003]

KING, R. 2002. Documentation for the GAMIT GPS Analysis Software – Data Analysis: Theory. [Internet:] ftp://bowie.mit.edu/pub/gps/updates/documentation/PDF_files/chap2.pdf. [Date of access: 10 January 2003]

KONDO, T., KURIHARA, N., KOYAMA, Y., SEKIDO, M., ICHIKAWA, R. 1998. Evaluation of repeatability of baseline lengths in the VLBI network around the Tokyo metropolitan area. *Geophysical Research Letters.* **25(7)**:1047–1050.

KUEMMEL, B. 1997. Temperature, humidity and dew point. [Internet:] http://www.faqs.org/faqs/meteorology/temp-dewpoint/. [Date of access: 17 May 2005]

LEICK, A. 1990. GPS Satellite Surveying. New York: Wiley-Interscience. 352 pp.

MEDEIROS, A.L. 2000. Radio propagation through the ionosphere layers. [Internet:] http://www.geocities.com/electricaltechnology/papers/_radiopro.html. [Date of access: 3 October 2003]

MIYAZAKI, S., IWABUCHI, T., HEKI, K., NAITO, I. 2003. An impact of estimating tropospheric delay gradients on precise positioning in the summer using the Japanese nationwide GPS array. *Journal of Geophysical Research.* **108**(B7):2335.

NCEP. 2006. NCEP/NCAR Reanalysis at CDC. [Internet:] http://www.cdc.noaa.gov/cdc/reanalysis/reanalysis.shtml. [Date of access: 7 January 2006]

NEUMEYER, J., BARTHELMES, F., COMBRINCK, L., DIERKS, O., FOURIE, P. 2002. Analysis Results from the SG Registration with the Dual Sphere

Superconducting Gravimeter at SAGOS (South Africa). *Bulletin d' Informations Marées Terrestres.* **135**:10607–10616.

NEUMEYER, J., DEL PINO, J., DIERKS, O., SUN, H.-P., PFLUG, H. 2005. Improvement of ocean loading correction on gravity data with additional tide gauge measurements. *Journal of Geodynamics.* **40**:104–111.

NEUMEYER, J., FOURIE, P., BARTHELMES, F., DIERKS, O., PFLUG, H. 2006. High precision gravity measurements with the Dual Sphere Superconducting Gravimeter in Sutherland (South Africa). *South African Journal of Geology* (manuscript accepted).

NEUMEYER, J., HAGEDOORN, J., LEITLOFF, J., SCHMIDT, T. 2004. Gravity reduction with three-dimensional atmospheric pressure data for precise ground gravity measurements. *Journal of Geodynamics.* **38**:437–450.

NIELL, A.E. 1996. Global mapping functions for the atmosphere delay at radio wavelengths. *Journal of Geophysical Research.* **101**:3227–3246.

NIELL, A.E, COSTER, A.J., SOLHEIM, F.S., MENDES, V.B., TOOR, P.C. LANGLEY, R.B., UPHAM, C.A. 2001. Comparison of measurements of atmospheric wet delay by radiosonde, water vapor radiometer, GPS and VLBI. *Journal of Atmospheric and Oceanic Technology.* **18**:830–850.

PACIONE, R., FIONDA, E., FERRARA, R., LANOTTE, R., SCIARRETTA, C., VESPE, F. 2002. Comparison of atmospheric parameters derived from GPS, VLBI and ground-based microwave radiometer in Italy. *Physics and Chemistry of the Earth.* **27**:309–316.

PIELKE, R.A. Sr. 2005. Land Use and Climate Change. *Science.* **310**:1625–1626.

POTTIAUX, E., WARNANT, R. 2002. First experiences with a water vapor radiometer at the Royal Observatory of Belgium. In *Proceedings of the 2002 Symposium of the IAG Sub-commission for Europe.* 426pp.

RIEPL, S., SCHLÜTER, W. 2000. Normal Point Algorithm for Reduction of Two Colour SLR Observations. [Internet:] http://www.wettzell.ifag.de/publ/2cnpt/nizza.html. [Date of access: 25 September 2003]

ROBERTSON, E. 2003. Atmosphere profile. [Internet:] http://faculty.millikin.edu/~erobertson.nsm.faculty.mu/Global_Studies_Climate2003/Atmprofile.html. [Date of access: 14 November 2003]

ROCKEN, C., SOKOLOSVSKIY, S., HUNT, D., JOHNSON, J.M. 2001. Improved mapping of tropospheric delays. *Journal of Atmospheric and Oceanic Technology*. **18**:1205–1213.

ROS, E., MARCIADE, J.M., GUIRADO, J.C., SARDÓN, E., SHAPIRO, I.I. 2000. A GPS-based method to model the plasma effects in VLBI observations. *Astronomy and Astrophysics*. **356**:357–362.

SAASTAMOINEN, J. 1972. Atmospheric correction for the troposphere and stratosphere in radio ranging of satellites. *The Use of Artificial Satellites for Geodesy*. **15**:247–251.

SCHERNECK, H.-G., BOS, M.S. 2002. Ocean Tide and Atmospheric Loading. *International VLBI Service for Geodesy and Astrometry: General Meeting Proceedings*.

SCHUELER, T., EISSFELLER, B., IRSIGLER, M., RODRÍGUEZ, J.A. 2004. Impact of Galileo on Geodynamics. *Unpublished proceedings of the 2004 WEGENER Symposium on Integrated Modelling of Crustal Deformation*.

SCHULZE, R. 2005. The Current Hydroclimatic Landscape in Southern Africa. *Climate Change and Water Resources in Southern Africa*. **6**:83–94. Ed. Schulze, R.

SOMIESKI, A. 2005. Geodetic Mobile Solar Spectrometer. Swiss Federal Institute of Technology, Zürich: Ph.D. Thesis. 224pp.

SOVERS, O.J., FANSELOW, J.L., JACOBS, C.S. 1998. Astrometry and geodesy with radio interferometry: experiments, models, results. *Reviews of Modern Physics*. **70(4)**:1393–1454.

SPILKER, J.J. Jr. 1994. Tropospheric Effects on GPS. *Global Positioning System: Theory and Applications*. **I**:517–546. Eds. Parkinson, B.W., Spilker, J.J. Jr. Reston: AIAA.

TREGONING, P., BOERS, R., O'BRIEN, D., HENDY, M. 1998. Accuracy of absolute precipitable water vapor estimates from GPS observations. *Journal of Geophysical Research.* **103**:28701–28710.

USWCL. 1999. U.S. Water Conservation Laboratory: Reference details for the Relative Humidity Equations. [Internet:] http://igscb.jpl.nasa.gov/components/prods.html. [Date of access: 17 May 2005]

VEY, S., CALAIS, E., LLUBES, M., FLORSCH, N., WOPPELMANN, G., HINDERER, J., AMALVICT, M., LALANCETTE, M.F., SIMON, B., DUQUENNE, F., HAASE, J.S. 2002. GPS measurements of ocean loading and its impact on zenith tropospheric delay estimates: a case study in Brittany, France. *Journal of Geodesy.* **76**:419–427.

WEBB, F., ZUMBERGE, J. 1995. An introduction to GIPSY/OASIS-II. Technical Report JPL D-11088. California Institute of Technology: Pasadena.

WONNACOTT, R.T., MERRY, C.L. 2006. The use of GPS for the estimation of precipitable water vapour for weather forecasting and monitoring in South Africa. *Survey Review* (manuscript accepted).

ZUMBERGE, J., HEFLIN, M., JEFFERSON, D., WATKINS, M., WEBB, F. 1997. Precise Point Positioning for the efficient and robust analysis of GPS data from large networks. *Journal of Geophysical Research.* **102**:5005–5017.

Appendix A GROUP AND PHASE VELOCITIES

To clarify the concepts of group and phase velocities discussed in Chapter 3, the argument given by Kaplan (1996:246–247) is followed. Consider two components S_1 and S_2 of an electromagnetic wave travelling in the x-direction, with frequencies ω_1 and ω_2, and phase velocities v_1 and v_2. The sum of these signals is

$$S = S_1 + S_2 = \sin\omega_1\left(t - \frac{x}{v_1}\right) + \sin\omega_2\left(t - \frac{x}{v_2}\right) \tag{A.1}$$

Using the trigonometric identity $\sin\alpha + \sin\beta = 2\cos\tfrac{1}{2}(\alpha-\beta)\cdot\sin\tfrac{1}{2}(\alpha+\beta)$ we find that (A.1) becomes

$$S = 2\cos\frac{1}{2}(\omega_1-\omega_2)\left[t - x\Big/\left(\frac{\omega_1-\omega_2}{\frac{\omega_1}{v_1}-\frac{\omega_2}{v_2}}\right)\right]\times\sin\left[\frac{1}{2}(\omega_1+\omega_2)t - \frac{1}{2}\left(\frac{\omega_1}{v_1}+\frac{\omega_2}{v_2}\right)x\right]$$

The cosine term refers to the wave group (the modulated sinusoid), which carries the information, and moves with velocity

$$v_g = \frac{\omega_1-\omega_2}{\frac{\omega_1}{v_1}-\frac{\omega_2}{v_2}} = \frac{\frac{v_1}{\lambda_1}-\frac{v_2}{\lambda_2}}{\frac{1}{\lambda_1}-\frac{1}{\lambda_2}} = \frac{\frac{v_1}{\lambda_1}-\frac{v_1}{\lambda_2}+\frac{v_1}{\lambda_2}-\frac{v_2}{\lambda_2}}{\frac{1}{\lambda_1}-\frac{1}{\lambda_2}} = v_1 - \lambda_1\frac{v_2-v_1}{\lambda_2-\lambda_1} \tag{A.2}$$

where λ_1 and λ_2 are the corresponding wavelengths.

For narrow bandwidths, such as found in GPS signals, the term $v_2 - v_1$ can be replaced by the differential dv, $\lambda_2 - \lambda_1$ by the differential $d\lambda$ and λ_1 by λ. Adding the subscript p to v to explicitly denote phase velocity, we obtain

$$v_g = v_p - \lambda\frac{dv_p}{d\lambda} \tag{A.3}$$

or equivalently

$$n_g = n_p + f\frac{dn_p}{df} \tag{A.4}$$

which implies that the difference between the group and phase velocities (or indices of refraction) depends on both the wavelength (or frequency) and the rate of change of phase velocity (or phase index of refraction) with wavelength (or frequency).

Expression (A.4) was derived from (A.3) by denoting the signal frequency by f and defining the indices of refraction as

$$n_p \equiv \frac{c}{v_p} \text{ and } n_g \equiv \frac{c}{v_g} \tag{A.5}$$

In a non-dispersive medium, by definition, wave propagation is independent of frequency and the signal phase and signal information (*i.e.* group) propagate at the same speed, *i.e.* $v_g = v_p$ and $n_g = n_p$ (Kaplan, 1996:247).

From equations (3.2) and (C.13) it is evident that the ionosphere and troposphere are dispersive and non-dispersive media respectively. This explains the simplified model of the atmosphere used in Chapter 3, in which the atmosphere is divided in only two parts (ionosphere and troposphere) with different properties for electromagnetic wave propagation.

Appendix B IONOSPHERIC DELAY FROM FIRST PRINCIPLES

The ionospheric delay experienced by radio waves is derived in this appendix, from first physical principles. The derivation serves to illuminate the fact that the dispersive nature of the ionosphere, which is often simply accepted as a given, can be derived from the most basic laws of electrodynamics. The ionosphere's properties are also of interest to this specific research project, because it is used to correct the GPS ranges on which GPS positioning is based.

We derive the refractivity of the ionosphere, following Chen (1984:114-116) and Choudhuri (1998:239-243), neglecting the effects of electron collisions and the Earth's magnetic field on wave propagation.

Free electrons in the ionosphere interact with electromagnetic waves travelling through the ionosphere. Maxwell's equations state

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{B.1}$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}. \tag{B.2}$$

$\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic components of the electromagnetic waves respectively, and $\mathbf{J}$ is the resulting current density.

The curl of (B.1) is

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla . \mathbf{E}) - \nabla^2 \mathbf{E} = -\nabla \times \frac{\partial \mathbf{B}}{\partial t}, \tag{B.3}$$

while the time derivative of (B.2) is

$$\nabla \times \frac{\partial \mathbf{B}}{\partial t} = \mu_0 \frac{\partial \mathbf{J}}{\partial t} + \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{B.4}$$

From (B.3) and (B.4) it follows that

$$\nabla^2\mathbf{E} - \nabla(\nabla.\mathbf{E}) = \mu_0 \frac{\partial \mathbf{J}}{\partial t} + \mu_0 \varepsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{B.5}$$

Assuming that the electromagnetic waves have an $e^{i(\mathbf{k}.\mathbf{r}-\omega t)}$ dependence, it follows from (B.5) that

$$k^2\mathbf{E} - \mathbf{k}(\mathbf{k}.\mathbf{E}) = i\omega\mu_0\mathbf{J} + \omega^2\mu_0\varepsilon_0\mathbf{E}. \tag{B.6}$$

Electromagnetic waves are transverse in nature, so that $\mathbf{k}.\mathbf{E} = 0$, and (B.6) becomes

$$(\omega^2 - c^2k^2)\mathbf{E} = \frac{-i\omega}{\varepsilon_0}\mathbf{J}. \tag{B.7}$$

Electromagnetic waves in both the radio and optical regions have such high frequencies that the ions can be considered as fixed; comparing their inertial mass to those of electrons, we expect that the motion of electrons will be the only source of current:

$$\mathbf{J} = -n_0 e\mathbf{v}, \tag{B.8}$$

with $-e$ the charge of an electron, n_0 the number density of electrons and $\mathbf{v}$ their velocity.

Newton's second law states $m\frac{d\mathbf{v}}{dt} = -e\mathbf{E}$; from the assumption that the waves have an $e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$ dependence, follows $\mathbf{v} = \frac{e\mathbf{E}}{im\omega}$. Combining (B.7) and (B.8), it follows that

$$(\omega^2 - c^2k^2)\mathbf{E} = \frac{n_0e^2}{\varepsilon_0 m}\mathbf{E}. \tag{B.9}$$

The plasma frequency is defined as $\omega_p^2 \equiv \frac{n_0e^2}{\varepsilon_0 m}$. From (B.9) we then obtain the dispersion relation

$$\omega^2 = \omega_p^2 + c^2k^2 \tag{B.10}$$

One can also write this as $ck = \omega\sqrt{1+\left(\frac{\omega_p}{\omega}\right)^2}$. From (B.10) and an alternative definition of the phase velocity, $v_p \equiv \frac{\omega}{k}$, follows $v_p^2 = c^2 + \omega_p^2/k^2 > c^2$; *i.e.* the phase velocity of an electromagnetic wave in the ionosphere is greater than c. The alternative definition of group velocity, $v_g \equiv \frac{\partial\omega}{\partial k}$, and (B.10) leads to $v_g = \frac{kc^2}{\omega} = \frac{c^2}{v_p} < c$, showing that the group velocity of an electromagnetic wave in the ionosphere is less than c. The effects of the ionosphere on microwaves are therefore referred to as phase advance and group delay.

Per definition the relationship between ω and k is given by $\omega(k) = \frac{ck}{n(k)}$. Group velocity can then also be expressed in the general form $v_g = \frac{\partial}{\partial k}\left(\frac{ck}{n(k)}\right) = \frac{c}{n(\omega)+\omega(dn/d\omega)}$, where $n(\omega)$ is the conventional index of refraction as depicted in Figure B in Appendix C. Furthermore, from Figure B it is evident that $(dn/d\omega) > 0$ and $n > 1$ for normal dispersion, while $(dn/d\omega)$ can also become large and negative in the vicinity of a resonance. The ionosphere is transparent in both the optical and radio frequencies, so that one would expect normal dispersion at these frequencies because of the absence of resonances. This argument also confirms that the group velocity of optical and radio waves in the ionosphere should be less than c and that the group refractivity should be positive.

From (A.5) and (B.10) the ionospheric indices of refraction can be defined as

$$n = \sqrt{1 \mp \left(\frac{\omega_p}{\omega}\right)^2} \qquad \text{(B.11)}$$

where the negative sign applies to phase velocities and the positive sign to group velocities. From (B.11) it is clear that $n_p = 0$ at $\omega = \omega_p$; therefore, k becomes zero and the wavelength of the electromagnetic wave becomes infinitely long in this limit. Thus,

ω_p is a cut-off frequency, so that no electromagnetic waves below this frequency can propagate through the ionosphere.

For the radio frequencies used for GPS communication (~1.2 and ~1.6 GHz) and the optical frequencies used for satellite laser ranging (2.8×10^{14} Hz to 4.3×10^{14} Hz), $\frac{\omega_p}{\omega} << 1$ and $\frac{\omega_p}{\omega} <<< 1$ respectively, and by means of the binomial expansion $\sqrt{1 \pm \varepsilon} \cong 1 \pm \frac{1}{2}\varepsilon$, (B.11) becomes

$$n \cong 1 \mp \frac{1}{2}\frac{\omega_p^2}{\omega^2} \text{ (radio) and } n \cong 1 \text{ (optical).} \tag{B.12}$$

From (B.12) one can determine that the path correction is proportional to the total electron content (TEC), and inversely proportional to the square of the frequency of the incoming electromagnetic wave. It is also apparent from (B.12) that the transport of optical waves in the ionosphere is non-dispersive, while the transport of radio waves in the ionosphere is dispersive.

Therefore, the first-order approximation of the ionospheric index of refraction can be written as

$$n = 1 \mp \frac{40.3 \text{ Hz}^2\text{m}^3}{f^2} n_0 \tag{B.13}$$

for the case of radio waves. In the ionosphere n_0 is typically of the order of $10^{12}\,\text{m}^{-3}$, so that the indices of refraction of a 1.6 GHz GPS signal in the ionosphere are typically $n_p \simeq 0.999984$ and $n_g \simeq 1.000016$.

The measured range between a satellite and a receiver is

$$S = \int_{SV}^{Receiver} n ds \tag{B.14}$$

whereas the true line-of-sight (*i.e.* geometric) range is

$$l = \int_{SV}^{Receiver} dl \tag{B.15}$$

The path length difference due to ionospheric refraction is therefore

$$\Delta S_{iono} = S - l = \int_{SV}^{Receiver} \left(1 \mp \frac{40.3\ \mathrm{Hz^2 m^3}}{f^2} n_0\right) ds - \int_{SV}^{Receiver} dl \tag{B.16}$$

The delay will be small compared to the satellite-to-user distance (Kaplan, 1996:249), so that the first term in (B.16) can be integrated along the line-of-sight path, yielding

$$\Delta S_{iono} = \mp \frac{40.3\ \mathrm{Hz^2 m^3}}{f^2} \int_{SV}^{Receiver} n_0 dl \tag{B.17}$$

Therefore, equation (B.17) gives the additional ionospheric path length, as used in Section 3.1.

Appendix C TROPOSPHERIC REFRACTIVE INDEX

In this appendix the refractive index of the troposphere, a dielectric medium, is derived from first physics principles. Furthermore, the non-dispersive nature of the troposphere is proven.

We make a model of dispersion that takes place on atomic scale in a dielectric medium, such as the troposphere, which is the lower, electrically neutral part of the atmosphere. We follow the derivation by Griffiths (1999:398-404), which is only an approximation of the quantum mechanical model, but yields satisfactory results nevertheless.

The electrons in a dielectric substance are bound to specific molecules. We shall picture each electron as attached to the end of an imaginary spring, with force constant k_{spring}, so that $F_{binding} = -k_{spring}x = -m\omega_0^2 x$ is the centrifugal force acting on the electron as it orbits the atom's nucleus, with x being the displacement from equilibrium, m the mass of the electron and $\omega_0 = \sqrt{k_{spring}/m}$ the natural oscillation frequency, as shown in Figure A.

The electrons will also experience a damping force $F_{damping} = -m\gamma(dx/dt)$, with γ a damping coefficient and where the negative sign follows from the fact that the force is in the direction opposite to the velocity of the electron. This damping is a result of the fact that oscillating charges radiate electromagnetic waves, carrying off energy with them.

In the presence of an electromagnetic wave of frequency ω, polarised in the x-direction, there is a driving force $F_{driving} = qE = qE_0\cos(\omega t)$ acting on the electron, with q the charge of the electron and E_0 the amplitude of the wave.

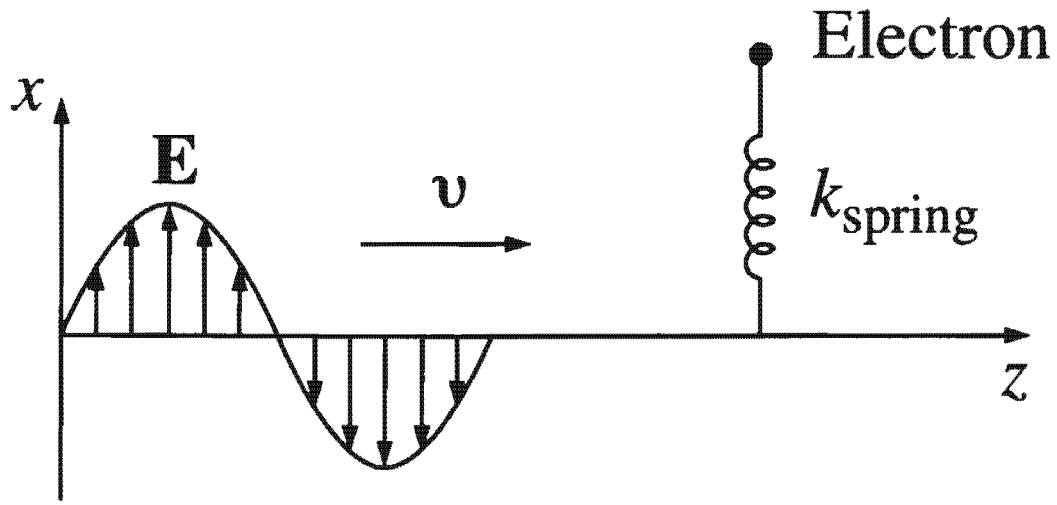

Figure A: The electrons in a dielectric substance are pictured as if attached to the end of an imaginary spring, and driven by a varying electric field (Griffiths, 1999:400).

Newton's second law now states

$$m\frac{d^2x}{dt^2} = F_{total} = F_{binding} + F_{damping} + F_{driving}$$

or

$$m\frac{d^2x}{dt^2} + m\gamma\frac{dx}{dt} + m\omega_0^2 x = qE_0\cos(\omega t). \qquad \text{(C.1)}$$

This model describes the electron as a damped harmonic oscillator, driven at frequency ω.

Consider (C.1) as the real part of the complex equation

$$\frac{d^2\tilde{x}}{dt^2} + \gamma\frac{d\tilde{x}}{dt} + \omega_0^2\tilde{x} = \frac{q}{m}E_0 e^{-i\omega t}. \qquad \text{(C.2)}$$

In the steady state, the system oscillates at the driving frequency ω, *i.e.* $\tilde{x}(t) = \tilde{x}_0 e^{-i\omega t}$. Placing this in (C.2), it follows that $\tilde{x}_0 = \dfrac{q/m}{\omega_0^2 - \omega^2 - i\gamma\omega}E_0$.

The dipole moment (defined as $\mathbf{p} = q\mathbf{x}$) is the real part of

$$\tilde{p}(t) = q\tilde{x}(t) = \frac{q^2/m}{\omega_0^2 - \omega^2 - i\gamma\omega} E_0 e^{-i\omega t}. \tag{C.3}$$

Electrons in the same molecule that have different orientations will, of course, experience different damping coefficients and natural frequencies. Assume that there are f_j electrons with frequency ω_j and damping coefficient γ_j in each molecule. If there are N molecules per unit volume, the polarisation (defined as dipole moment (C.3) per unit volume) will be given by the real part of

$$\tilde{\mathbf{P}} = \frac{Nq^2}{m}\left(\sum_j \frac{f_j}{\omega_0^2 - \omega^2 - i\gamma_j\omega}\right)\tilde{\mathbf{E}}. \tag{C.4}$$

Because of the phase difference between $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{E}}$ we have to define a complex susceptibility by $\tilde{\mathbf{P}} = \varepsilon_0 \tilde{\chi}_e \tilde{\mathbf{E}}$.

The permittivity of the substance is defined as $\varepsilon = \varepsilon_0(1 + \chi_e)$, while the dielectric constant is defined as $\varepsilon_r = \varepsilon/\varepsilon_0$. For this model it implies that

$$\tilde{\varepsilon} = \varepsilon_0(1 + \tilde{\chi}_e)$$

and

$$\tilde{\varepsilon}_r = 1 + \frac{Nq^2}{m\varepsilon_0}\sum_j \frac{f_j}{\omega_0^2 - \omega^2 - i\gamma_j\omega}. \tag{C.5}$$

In a dispersive medium the wave equation for a given frequency is

$$\nabla^2\tilde{\mathbf{E}} = \tilde{\varepsilon}\mu_0 \frac{\partial^2\tilde{\mathbf{E}}}{\partial t^2},$$

with μ_0 the permeability of vacuum. It admits plane wave solutions of the form

$$\tilde{\mathbf{E}}(z,t) = \tilde{\mathbf{E}}_0 e^{i(\tilde{k}z - \omega t)}, \tag{C.6}$$

with complex wave number

$$\tilde{k} \equiv \sqrt{\tilde{\varepsilon}\mu_0}\,\omega. \tag{C.7}$$

We can also write the complex wave number in terms of real and imaginary components:

$$\tilde{k} = k + i\kappa. \tag{C.8}$$

Using (C.8), expression (C.6) now becomes $\tilde{\mathbf{E}}(z,t) = \tilde{\mathbf{E}}_0 e^{-\kappa z} e^{i(kz-\omega t)}$, indicating that the wave is being damped.

The power of the electromagnetic wave is proportional to E^2, and therefore also to $e^{-2\kappa z}$. Consequently, the quantity

$$\alpha \equiv 2\kappa \tag{C.9}$$

is called the absorption coefficient. Furthermore, the wave velocity $v = \omega/k$, and the index of refraction is

$$n = ck/\omega \tag{C.10}$$

For gasses, as in the case of the troposphere, the second term of the complex dielectric constant is small. Using equation (C.5), expression (C.7) and the binomial expansion $\sqrt{1+x} \cong 1+\tfrac{1}{2}x$ for small x, we can write

$$\tilde{k} = \frac{\omega}{c}\sqrt{\tilde{\varepsilon}_r} \cong \frac{\omega}{c}\left[1 + \frac{Nq^2}{2m\varepsilon_0}\sum_j \frac{f_j}{\omega_j^2 - \omega^2 - i\gamma_j\omega}\right],$$

so that it follows from (C.8) and (C.10) that

$$n = \frac{ck}{\omega} \cong 1 + \frac{Nq^2}{2m\varepsilon_0}\sum_j \frac{f_j(\omega_j^2 - \omega^2)}{(\omega_j^2 - \omega^2)^2 + \gamma_j^2\omega^2} \tag{C.11}$$

and from (C.9) follows

$$\alpha = 2\kappa \cong \frac{Nq^2\omega}{m\varepsilon_0 c}\sum_j \frac{f_j\gamma_j}{(\omega_j^2 - \omega^2)^2 + \gamma_j^2\omega^2}.$$

The index of refraction rises gradually with increasing frequency, except in the vicinity of a resonance $\omega = \omega_j$, where it drops sharply, as shown in Figure B. The resonances are caused by electrons being driven at specific (resonance) frequencies,

which correspond to relatively large amplitudes and, consequently, the dissipation of large amounts of energy due to the damping mechanism. Therefore, the resonances also coincide with frequencies of maximum absorption. This atypical behaviour in the vicinity of a resonance is called anomalous dispersion.

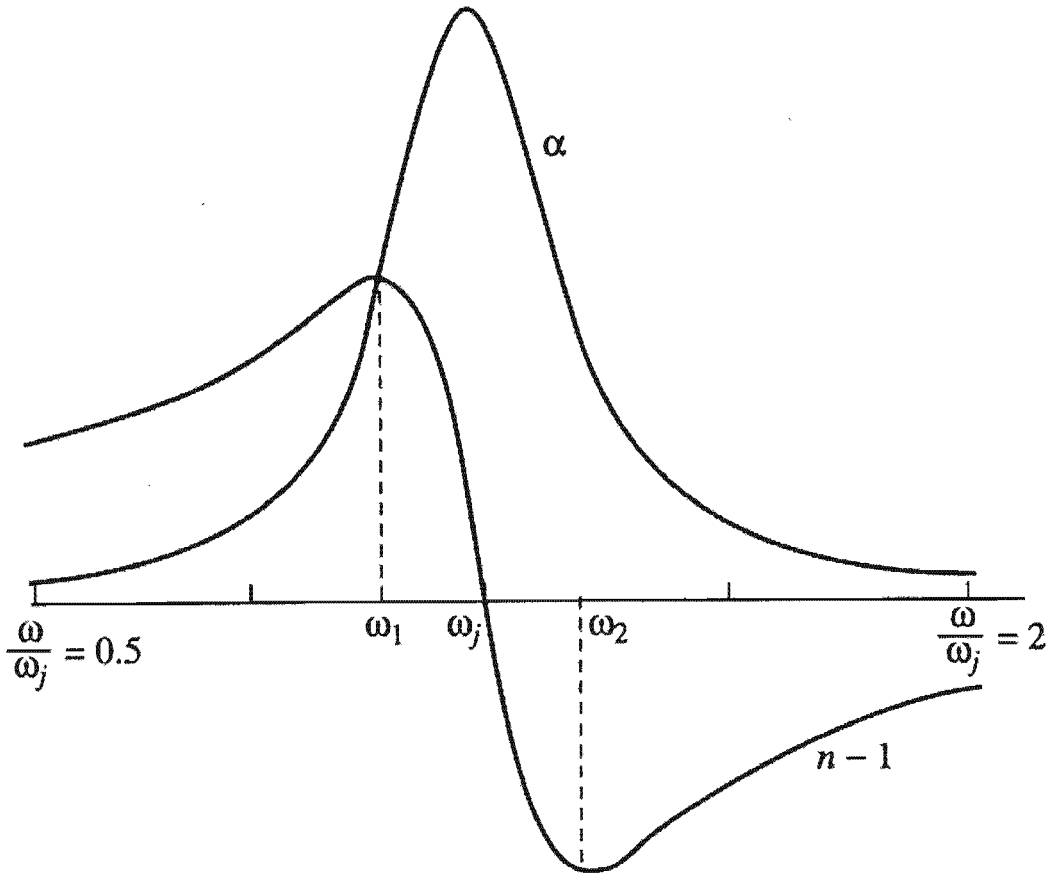

Figure B: Anomalous dispersion and absorption in the frequency region of a resonance (Griffiths, 1999:403).

In this study we consider radio and optical waves. In the troposphere we find that the natural frequencies, ω_j, typically occur at ultraviolet and higher frequencies, and also between the deep infrared and submillimetre wavelength-level, which explains why the atmosphere appears to be opaque at these frequencies. Therefore, in the case of the

optical and radio frequencies (with the notable exception of the absorption line of water vapour at 22 GHz) we can ignore damping and reduce (C.11) to

$$n = 1 + \frac{Nq^2}{2m\varepsilon_0}\sum_j \frac{f_j}{\omega_j^2 - \omega^2}. \tag{C.12}$$

Since $\omega << \omega_j$, we now write

$$\frac{1}{\omega_j^2 - \omega^2} = \frac{1}{\omega_j^2}\left(1 - \frac{\omega^2}{\omega_j^2}\right)^{-1} \cong \frac{1}{\omega_j^2}\left(1 + \frac{\omega^2}{\omega_j^2}\right),$$

so that (C.12) can be written as

$$n = 1 + \left(\frac{Nq^2}{2m\varepsilon_0}\sum_j \frac{f_j}{\omega_j^2}\right) + \omega^2\left(\frac{Nq^2}{2m\varepsilon_0}\sum_j \frac{f_j}{\omega_j^4}\right).$$

We can also write this in terms of wavelength ($\lambda = 2\pi c/\omega$):

$$n = 1 + N_n\left(1 + \frac{B}{\lambda^2}\right). \tag{C.13}$$

This is known as the Cauchy equation, with N_n the refractivity and B the dispersion coefficient. B was found to be $1.7\times10^{-14}\,\mathrm{m}^2$ for the atmosphere (Riepl & Schlüter, 2000).

Optical wavelengths are typically of the order of $\sim 5.0\times10^{-7}\,\mathrm{m}$, so that λ^2 is comparable to B; N_n is typically of the order of ~0.4 for air. Cauchy's equation (C.13) applies reasonably well to most gases in the optical region (Griffiths, 1999:404). One can therefore conclude that the transport of optical waves in the troposphere is dispersive.

The typical radio wavelengths we consider are of the order of ~0.2 m (L-band), so that $\frac{B}{\lambda^2} << 1$, and consequently $N_n \approx n - 1$. Thus we can conclude that the propagation of radio waves in the troposphere is non-dispersive.

In the troposphere $n > 1$ for both the radio and optical cases; from the definition of the index of refraction follows that an electromagnetic wave traversing the atmosphere will experience a delay (or longer apparent path length) relative to the propagation time

(or apparent path length) of the same wave in vacuum, if the propagation velocity is assumed to be c in both cases.

The index of refraction for radio waves in the troposphere is always greater than 1 (per definition equals 1 for vacuum), as can be seen from expression (C.13). These waves traversing the troposphere experience a resulting delay, which appears to add up to ~2 metres to the path length of the waves; this is $\sim 10^{-5}\%$ of the total path length of 20 200 km.

In conclusion, the non-dispersive index of refraction derived in this appendix for the propagation of radio waves in the troposphere, can be used in equation (3.12) when deriving the tropospheric delay of GPS signals, as in Section 3.2.

Appendix D EMPIRICAL MODELS AND MAPPING FUNCTIONS

The interested reader is referred to Spilker's (1994:534–544) discussion of other empirical models not described in this appendix, namely the Black and Eisner model, the Berman water vapour zenith delay model, the Altshuler and Kalaghan delay model, ray tracing models, simplified models, the Lanyi mapping function and the GPS Control Segment delay estimate.

Saastamoinen total delay model

Spilker (1994:534) states that Saastamoinen presented one of the first models of refraction of the troposphere that estimates delay as a function of elevation angle. Saastamoinen's (1972:247–251) standard model for the tropospheric delay correction for radio frequency ranging, for elevation angles $\theta \geq 10°$, is given by

$$\Delta L = 0.002277\,(1+D)\sec\psi\left[P+\left(\frac{1255}{T}+0.005\right)e - B\tan^2\psi\right]+\delta$$

where ΔL is the delay correction in metres, P and e are the total and water vapour pressures respectively in millibars, T is the temperature in Kelvin, $\psi = 90° - \theta$ is the zenith angle and B and δ are height-dependent correction terms given in Spilker (1994:535). The value of D is $0.00266\times\cos 2\lambda - 2.8\times10^{-7}H$, with λ the latitude and H the height (in metres) above the ellipsoid.

Hopfield two-quartic model

Hopfield has developed a dual quartic zenith model of refraction, different fourth-order polynomials for the dry and wet atmospheric profiles (Spilker, 1994:534). The two-quartic model of refractivity as a function of altitude h is given by

$$N_i(h) = N_{i0}\left(1-\frac{h}{h_i}\right)^4 \quad \text{for } h \le h_i$$

where $i = d, w$ for the hydrostatic and wet components respectively, N_{i0} are the respective refractivities at the surface and $h_d = 43$ km, $h_i = 12$ km. Therefore, to obtain the delay, integration along the ray path must be done up to 12 km for the wet component for the troposphere (where almost all atmospheric water vapour occur) and up to 43 km for the dry atmosphere contribution from the troposphere and stratosphere:

$$\Delta L = 10^{-6}\sum_i \int_0^{h_i} N_{i0}\left(1-\frac{h}{h_i}\right)^4 dh = \frac{10^{-6}}{5}\left[N_{d0}h_d + N_{w0}h_w\right]$$

Davis, Chao and Marini mapping functions

As mentioned earlier, the mapping function attempts to relate actual path delays to zenith delays as a function of elevation angle; these mapping functions can be used conjunctly with models for the zenith delay (*e.g.* the Saastamoinen or Hopfield models). Near zenith a simple sine mapping function is sufficient, but at lower elevation angles the troposphere's non-uniform and finite width spherical shell make this simple mapping function inadequate. The following brief overview of three widely used mapping functions is a summary of a discussion thereof by Spilker (1994:538–539).

Marini presented a continued fraction version of the mapping function:

$$m(\theta) = \cfrac{1}{\sin\theta + \cfrac{a}{\sin\theta + \cfrac{b}{\sin\theta + \cfrac{c}{\sin\theta + \ldots}}}}$$

where a, b, c,... are constants. Chao's mapping functions (one for the hydrostatic and one for the wet component) have a similar form, but use only the first two terms and substitute the second $\sin\theta$ for $\tan\theta$ to make $m(90°) = 1$. These mapping functions are

$$m_i(\theta) = \frac{1}{\sin\theta + \frac{a_i}{\tan\theta + b_i}}$$

where $a_d = 0.00143$, $b_d = 0.0445$, $a_w = 0.00035$, $b_w = 0.017$.

The so-called "Davis Cfa 2.2" mapping function is more sophisticated, wherein the coefficients a, b, and c are dependent on surface pressures, temperature, lapse rates and tropospheric height. The Davis mapping function is of the form

$$m(\theta) = \frac{1}{\sin\theta + \frac{a}{\tan\theta + \frac{b}{\tan\theta + c}}}$$

Niell mapping function

The Niell mapping function (NMF) proposed by Niell (1996:3230) is most commonly used in space geodesy, because of its accuracy at elevation angles from 90° down to below 10°, and it does not require any meteorological observations. NMF is of the form

$$m(\theta) = \frac{1 + \frac{a}{1 + \frac{b}{1 + c}}}{\sin\theta + \frac{a}{\sin\theta + \frac{b}{\sin\theta + c}}} \qquad \text{(D.1)}$$

with a, b and c different parameters for the hydrostatic and wet mapping functions, and also varying with latitude. Figure C shows the form of equation (D.1), where the values of the parameters as given by Niell (1996:3250) for latitudes of 45°, were used. This figure shows that the mapping function is nearly a power law in θ, approximately of the form $f = 28.9\theta^{-0.77}$ (shown by the straight line).

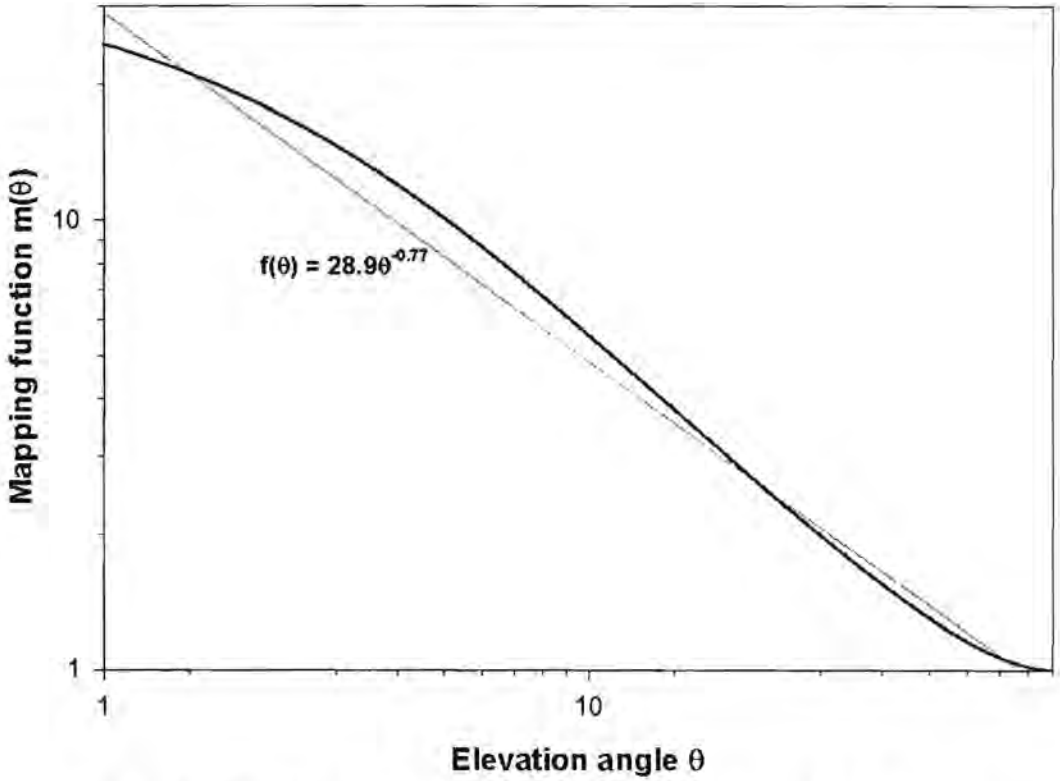

Figure C: The Niell Mapping Function (NMF) for latitudes of 45 degrees (Combrink, 2003:26).

Vienna mapping function

A brief mention of a more recently developed mapping function is made here. The use of the Vienna mapping function (VMF1), presented by Boehm & Schuh (2004), could significantly improve the precision of geodetic results such as baseline lengths and station heights and account for atmospheric phenomena. However, VMF1 requires numerical weather analysis information for the estimation of mapping parameters and was consequently not used in the presented study.

Appendix E DEGREE OF CORRELATION

Since the obtained correlation coefficient r cannot be used directly to indicate the degree of correlation, the probability $P_c(r,N)$ that a random sample of N uncorrelated experimental data points would yield a correlation coefficient as large as or larger than the observed $|r|$, are calculated by

$$P_c(r,N) = 2\int_{|r|}^{1} P_r(\rho,\nu)\,d\rho \tag{E.1}$$

where $\nu = N-2$ is the number of degrees of freedom and

$$P_r(r,\nu) = \frac{1}{\sqrt{\pi}}\frac{\Gamma\left[(\nu+1)/2\right]}{\Gamma(\nu/2)}\left(1-r^2\right)^{(\nu-2)/2} \tag{E.2}$$

with $\Gamma(n)$ the gamma function (Bevington, 1969:122–123).

Appendix F MULTIPLE REGRESSION WITH DUMMY VARIABLES

In this appendix a model for seasonal patterns is described, which uses regression with dummy variables. A seasonal model, with trend m_t and seasonal function s (with period p), is given by

$$X_t = m_t + s_t + Y_t,\ t = 1,\ldots,n \tag{F.1}$$

with Y_t the offsets from the model of the individual observations and $s_{t-p} = s_t$. Dummy variables $z_{i,t}$ are introduced in the structure of s_t as follows:

$$s_t = \sum_{i=1}^{p-1} \beta_i z_{i,t} \tag{F.2}$$

$$z_{i,t} = \begin{cases} 1 \text{ if } t = i \\ 0 \text{ otherwise} \end{cases} \tag{F.3}$$

The jth order trend is represented by $m_t = \sum_{i=0}^{j} \alpha_i t^i$. The seasonal parameter for season p is arbitrarily set to zero.

This model can also be written in matrix form:

$$\underline{X} = Z\underline{\gamma} + \underline{Y}, \text{ or} \tag{F.4}$$

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_{s-1} \\ X_s \\ X_{s+1} \\ \vdots \\ X_{n-1} \\ X_n \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 2 & \cdots & 2^j & 0 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 3 & \cdots & 3^j & 0 & 0 & 1 & \cdots & 0 & 0 \\ 1 & 4 & \cdots & 4^j & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 1 & s-1 & \cdots & (s-1)^j & 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & s & \cdots & s^j & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & s+1 & \cdots & (s+1)^j & 1 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 1 & n-1 & \cdots & (n-1)^j & 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & n & \cdots & n^j & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{j-1} \\ \alpha_j \\ \beta_1 \\ \beta_2 \\ \vdots \\ \vdots \\ \beta_{s-2} \\ \beta_{s-1} \end{bmatrix} + \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ \vdots \\ Y_{s-1} \\ Y_s \\ Y_{s+1} \\ \vdots \\ Y_{n-1} \\ Y_n \end{bmatrix} \tag{F.5}$$

Therefore, the least-squares solution of the parameter vector $\underline{\gamma}$ is given by

$$\hat{\underline{\gamma}} = \left(Z^T Z\right)^{-1} Z^T \underline{X} \tag{F.6}$$

Based on existing knowledge of the dataset, we may choose to assign different weights to each of the data points. Weights can be presented in an $n \times n$ diagonal matrix W, so that (F.6) can be rewritten as

$$\hat{\underline{\gamma}} = \left(Z^T W Z\right)^{-1} Z^T W \underline{X} \tag{F.7}$$

This equation will be employed, as described in Section 5.2.3, by the modified residual stacking (MRS) method to obtain multipath-free line-of-sight estimates of the tropospheric wet delay. The weighting function will be used to ensure that more recent data will carry more weight in the analysis than older data, as one would expect intuitively.

Appendix G DESCRIPTIVE QUANTITIES OF TIME-SERIES BEHAVIOUR

Consider an observed time series $x_1, x_2, ..., x_n$. Then

- the sample mean of the time series is defined as

$$\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t \tag{G.1}$$

- the sample autocovariance coefficient at lag k of the time series is defined as

$$c_k = \frac{1}{n}\sum_{t=1}^{n-k}\left(x_t - \bar{x}\right)\left(x_{t+k} - \bar{x}\right) \tag{G.2}$$

- the sample variance of the time series is defined as

$$c_0 = \frac{1}{n-1}\sum_{t=1}^{n}\left(x_t - \bar{x}\right)^2 \tag{G.3}$$

- and the sample autocorrelation coefficient at lag k of the time series is defined as

$$r_k = \frac{c_k}{c_0} \tag{G.4}$$

As a function of k we call c_k the sample autocovariance function (sACOVF) and r_k the sample autocorrelation function (sACF).

The sample partial autocorrelation function (sPACF) $\phi(k)$ is estimated by

$$\hat{\phi}(k) = \frac{\det\left(\widehat{P_k^*}\right)}{\det\left(\widehat{P_k}\right)} \tag{G.5}$$

where $\widehat{P_k} = \left\{\hat{p}_{ij}\right\}$ is the $k \times k$ autocorrelation matrix with $\hat{p}_{ij} = r_{|i-j|}$. $\widehat{P_k^*}$ is the same as $\widehat{P_k}$ but with the last column replaced with $\left(r_1, r_2, ..., r_k\right)^T$.

The quantities described in this appendix are descriptive quantities of time series behaviour, such as the behaviour of GPS processing residuals. Of particular interest in this research project is the sample autocorrelation function (sACF), which has been used in Section 5.2.3 to determine the correlation between the GPS processing residuals measured on days separated by a lag of k days.